

Col - 4102793-11-0335085

# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER



With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich  
Alan R. Anderson  
Kurt Baier  
Stephen F. Barker  
Monroe Beardsley  
Nuel D. Belnap, Jr.  
Myles Brand  
Robert Butts  
Roderick M. Chisholm  
L. Jonathan Cohen

James Collins  
Anthony Cua  
Arthur C. Danto  
Richard M. Gale  
Peter Thomas Geach  
Marjorie Grene  
Adolf Grünbaum  
R. M. Hare  
Rom Harré  
Carl G. Hempel

Raymond Klibansky  
Hugues Leblanc  
Gerald J. Massey  
George Mavrodes  
Richard Rorty  
George A. Schrader  
Wilfrid Sellars  
Ernest Sosa  
Richard Taylor  
John W. Yolton

VOLUME 11/NUMBER 1

JANUARY, 1974

## CONTENTS

|  |    |  |    |
|--|----|--|----|
| I. ROBERT AUDI: <i>Moral Responsibility, Freedom, and Compulsion</i> . . . . . | 1  | VI. BARBARA M. HUMPHRIES: <i>Knowing the Meaning</i> . . . . .   | 55 |
| II. MARSHALL SWAIN: <i>Epistemic Defeasibility</i> . . . . .                   | 15 | VII. THOMAS M. OLSHEWSKY: <i>The Analogical Argument for Knowledge of Other Minds Reconsidered</i> . . . . .           | 63 |
| III. ROBERT M. GORDON: <i>The Aboutness of Emotions</i> . . . . .              | 27 | VIII. DAVID M. JOHNSON: <i>The Temporal Dimension of Perceptual Experience: A Non-Traditional Empiricism</i> . . . . . | 71 |
| IV. A. F. MACKAY: <i>Evaluative Asymmetry</i> . . . . .                        | 37 | <i>Books Received</i> . . . . .  | 77 |
| V. BARRY MILLER: <i>Making Sense of "Necessary Existence"</i> . . . . .        | 47 |  |    |

# AMERICAN PHILOSOPHICAL QUARTERLY

W8

The *American Philosophical Quarterly* welcomes <sup>NO</sup> by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$14.00, increasing by \$1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$11.20, increasing by \$1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of \$6 per group of 50.

## SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and eighteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and five dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

## MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

335085



# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich  
Annette Baier  
Kurt Baier  
Stephen F. Barker  
Monroe Beardsley  
Nuel D. Belnap, Jr.  
Myles Brand  
Robert Butts  
Roderick M. Chisholm  
L. Jonathan Cohen

James Collins  
Antonio S. Cua  
Arthur C. Danto  
Richard M. Gale  
Peter Thomas Geach  
Marjorie Grene  
Adolf Grünbaum  
Rom Harré  
Carl G. Hempel

Hugues Leblanc  
Keith Lehrer  
Gerald J. Massey  
George Mavrodes  
Richard Rorty  
George A. Schrader  
Wilfrid Sellars  
Ernest Sosa  
Richard Taylor  
John W. Yolton



VOLUME 11 (1974)

---

PUBLISHED IN OXFORD BY BASIL BLACKWELL

# AMERICAN PHILOSOPHICAL QUARTERLY

## CONTENTS OF VOLUME 11 (1974)

|   |     |
|---|-----|
| ACHINSTEIN, PETER: <i>The Identity of Properties</i>  | 257 |
| ALDRICH, VIRGIL C.: <i>Sight and Light</i>  | 317 |
| AUDI, ROBERT: <i>Moral Responsibility, Freedom, and Compulsion</i>                                      | I   |
| BEVERSLUIS, JOHN: <i>Socratic Definition</i>  | 331 |
| CORRIGENDA  | 345 |
| CRAWFORD, DAN: <i>Bergmann on Perceiving, Sensing, and Appearing</i>                                    | 103 |
| CURLEY, EDWIN: <i>Recent Work on 17th Century Continental Philosophy</i>                                | 235 |
| FISHER, JOHN: <i>Universalizability and Judgments of Taste</i>  | 219 |
| GORDON, ROBERT M.: <i>The Aboutness of Emotions</i>   | 27  |
| GOTTLIEB, DALE: <i>Foundations of Logical Theory</i>  | 337 |
| HOFFMAN, ROBERT: <i>Vulgarity</i>   | 323 |
| HUMPHRIES, BARBARA M.: <i>Knowing the Meaning</i>   | 55  |
| JOHNSON, DAVID M.: <i>The Temporal Dimension of Perceptual Experience: A Non-Traditional Empiricism</i> | 71  |
| KELLENBERGER, J.: <i>God and Mystery</i>  | 93  |
| LEHRER, KEITH: <i>Truth, Evidence and Inference</i>   | 79  |
| LOCKE, DON: <i>Reasons, Wants, and Causes</i>   | 169 |
| MACKAY, A. F.: <i>Evaluative Asymmetry</i>  | 37  |
| MARCUS, RUTH BARGAN: <i>Classes, Collections, and Individuals</i>                                       | 227 |
| MATTHEWS, GARETH: <i>Paradoxical Statements</i>   | 133 |
| MILLER, BARRY: <i>Making Sense of "Necessary Existence"</i>   | 47  |
| OLSHESKY, THOMAS M.: <i>The Analogical Argument for Knowledge of Other Minds Reconsidered</i>           | 63  |
| PARENT, WILLIAM A.: <i>Some Recent Work on the Concept of Liberty</i>                                   | 149 |
| QUINN, WARREN S.: <i>Theories of Intrinsic Value</i>  | 123 |
| RIPLEY, CHARLES: <i>A Theory of Volition</i>  | 141 |
| SCOTT, STEPHEN: <i>Rituals and "Ought"</i>  | 211 |
| SCHACHT, RICHARD: <i>On "Existentialism," Existenz-Philosophy and Philosophical Anthropology</i>        | 291 |
| SOLOMON, ROBERT: <i>Hegel's Epistemology</i>  | 277 |
| SOSA, ERNEST: <i>The Concept of Knowledge. How Do You Know?</i>   | 113 |
| SWAIN, MARSHALL: <i>Epistemic Defeasibility</i>   | 15  |
| TEGHRARIAN, SOUREN: <i>Linguistic Rules and Semantic Interpretation</i>                                 | 307 |
| TREBILCOT, JOYCE: <i>Aprudentialism</i>   | 203 |
| YOUNG, J. MICHAEL: <i>The Ontological Argument and the Concept of Substance</i>                         | 181 |
| YOUNG, ROBERT: <i>Petitioning God</i>   | 193 |



# I. MORAL RESPONSIBILITY, FREEDOM, AND COMPULSION

ROBERT AUDI

PHILOSOPHERS have very often held that a person can be morally responsible for having done something only if he did it freely.<sup>1</sup> I shall argue that this view is false; but it will be clear that the notion of free action is still quite important for moral philosophy. A great deal has been written on the free will problem; yet until very recently most writers on the problem have been chiefly concerned with the issue of whether determinism, most often vaguely construed as the thesis that every event has a cause, is compatible with the existence of free action. In the English-speaking world, at least, philosophers have increasingly tended to defend the compatibility of determinism and free action. Some of them have accepted a kind of compatibilism on the ground that even if determinism is true, actions are immanently caused by the agent and hence need not be determined ultimately by events beyond his control;<sup>2</sup> other writers have held that even if all events are determined, actions are not mere events and are not caused;<sup>3</sup> and still other writers have argued that actions may be free even though they are events caused by other events.<sup>4</sup> But despite the extensive literature on freedom and determinism, too few

writers have tried to construct a detailed account of free action itself.<sup>5</sup>

My principal aim will be to give an account of free action and its relation to moral responsibility. There is a technical use of "free action," but we do sometimes ask whether someone did a particular thing freely, or (meaning approximately the same thing) of his own free will. These are the main uses of "free" I shall try to explicate. Surprising as it may seem, this task does not require discussion of what it is for the will to be free. One can, as Frankfurt does,<sup>6</sup> make sense of "free will," but this notion is only indirectly connected with that of free action. I shall devote little space to arguing directly that free actions may be "determined." But if my account is correct, it will serve as an indirect argument; for on my account an action's being free entails neither that it is nor that it is not determined.

My first question will be this: if a person,  $x$ , is morally responsible for having performed some action,  $A$ , at a given time,  $t$  ( $xMRAt$ ), does it follow that  $x$  did  $A$  freely at  $t$  ( $xFAt$ )? I shall then consider the notion of doing  $A$  freely, and finally the relation between moral responsibility and free action. I hope

<sup>1</sup> For instance, A. C. Ewing maintains that "responsibility requires freedom either in the indeterminist sense or at least in the determinist sense." See his *Ethics* (London, 1953), p. 145. The view that a person's being morally responsible for an action entails (or presupposes) that he performed it freely is also held by Kant in his *Foundations of the Metaphysics of Morals*, trans. by Lewis White Beck (New York, 1959), p. 80; C. A. Campbell, e.g., in "Is 'Freewill' a Pseudo-Problem?", *Mind*, vol. 60 (1951), reprinted in Bernard Berofsky (ed.), *Free Will and Determinism* (New York, 1966), p. 129; S. I. Benn and R. S. Peters, *The Principles of Political Thought* (London, 1959), pp. 229 and 241; Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy*, vol. 68 (1971), p. 19; P. H. Nowell-Smith: see the citation by Maurice Cranston in *Freedom: A New Analysis* (London, 1953), p. 92; and Moritz Schlick, who speaks of freedom as "the presupposition of moral responsibility." See his "When Is a Man Responsible?" from *The Problems of Ethics* (New York, 1939) and reprinted in Berofsky, *op. cit.*, p. 59. Cf. D. J. O'Connor, *Free Will* (New York, 1971), p. 30.

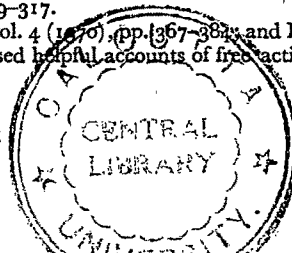
<sup>2</sup> See, e.g., Roderick M. Chisholm, "Freedom and Action," in Keith Lehrer (ed.), *Freedom and Determinism* (New York, 1966), pp. 11 and 17.

<sup>3</sup> For discussion of this view, see O'Connor, *op. cit.*, pp. 99-111. A leading proponent of it is A. I. Melden, *Free Action* (London, 1961).

<sup>4</sup> For a good statement of this view, see Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, New Jersey, 1970), esp. ch. Six. Another vigorous defense of compatibilism is offered by Adolf Grünbaum in "Free Will and the Laws of Human Behavior," *American Philosophical Quarterly*, vol. 8 (1971), pp. 299-317.

<sup>5</sup> Cranston, *op. cit.*; Gerald Dworkin, "Acting Freely," *Nous*, vol. 4 (1970), pp. 367-382; and Felix E. Oppenheim, *Dimensions of Freedom* (New York, 1961) are among those who have proposed helpful accounts of free action. But none of these accounts seems to me fully satisfactory or adequately detailed.

<sup>6</sup> See Frankfurt, *op. cit.*



to enhance our understanding of freedom as contrasted with compulsion and to show, through an account of both, how we can defeat certain popular but unwarranted disclaimers of moral responsibility.

### I. MORAL RESPONSIBILITY, FREE ACTION, AND AVOIDABILITY

We should first note some of the important kinds of moral responsibility. When sentences of the form " $x$  is morally responsible for doing  $A$ " have future reference, they are usually used to attribute to  $x$  an obligation to do  $A$ . Typically such obligations are of the kind Hart calls role responsibilities.<sup>7</sup> Our concern, however, is with cases in which the reference of " $x$  is morally responsible for doing  $A$ " is past. In these cases, there seems to be both a weak and strong sense of "moral responsibility." The weak and less common sense I shall call the accountability sense.<sup>8</sup> To say that  $x$  is morally responsible in this sense for doing  $A$  is to say that it would be appropriate to ask him to give a moral justification of his having done  $A$ . ("Just remember that you will be morally responsible for what you do about this.") In the stronger sense, which I shall call the liability sense,<sup>9</sup> to say that  $x$  is morally responsible for doing  $A$  is to say that he is *prima facie* liable to moral blame for doing it. It is chiefly, perhaps only, moral responsibility in this sense that is implicit in *holding* someone morally responsible. ("He has every right to hold  $x$  morally responsible for having done that.") There is often, though not always, a further implication, in ascriptions of liability responsibility, that  $x$  ought (morally) to be punished, or at least to pay compensation, for doing  $A$ . Perhaps in part because we rarely speak of people as morally responsible for having done  $A$  unless we think they may be blameworthy for doing it, some philosophers seem to have thought that being morally responsible for an action entails being blameworthy for it.<sup>10</sup> But this is surely a mistake. One may be morally responsible, without being blameworthy, for breaking certain promises and for giving one's twelve-year-old child permis-

sion to go on a potentially dangerous camping trip. Indeed, as these cases suggest, " $x$ MR*A*t" is compatible with  $A$ 's being praiseworthy. There is a fairly common philosophical use of " $x$ MR*A*t" in which it means roughly, "If  $A$  is wrong  $x$  is blameworthy for it, and if  $A$  is right  $x$ 's doing it is praiseworthy (or worthy of approval)." I shall not discuss this technical notion, but much of what emerges concerning moral responsibility will apply to the notion.

Both responsibility in the liability sense and responsibility in the accountability sense figure in contexts in which there is a question whether  $x$  was morally responsible *over* a given period. Hart calls this kind of responsibility capacity responsibility.<sup>11</sup> We should also note what might be called *dispositional responsibility*: this is our concern when we ask whether  $x$  is a morally responsible person. This paper should provide much of the basis for an account of capacity and dispositional responsibility, since both can be partially analyzed in terms of notions I shall be explicating. But I shall concentrate on liability responsibility, though my main points about this will also apply to responsibility in the weaker, accountability sense. Liability responsibility admits of degrees, but for our purposes it will be adequate simply to explore some important conditions for  $x$ 's bearing at least some degree of liability responsibility for doing  $A$ .

Let us first ask whether " $x$ MR*A*t" entails " $x$ FA*t*." Certainly in most cases in which  $x$  is morally responsible for having done  $A$  he did  $A$  freely. But suppose  $x$  fails to notice a flashing sign that says "Children Crossing," proceeds at substantially above the speed limit, and hits a child who is crossing. Even if  $x$  hit the child by accident, it is clear that  $x$  would be morally and legally responsible for hitting him. Yet it would be wrong to say either that he hit him freely or that his hitting him was a free action. One may object that  $x$  is responsible primarily for recklessness. But that is compatible with his also being morally responsible for hitting the child. It is also compatible with the possibility that his hitting the child is a consequence of his recklessness. For we are morally responsible

<sup>7</sup> H. L. A. Hart, "Responsibility and Retribution," in Hart's *Punishment and Responsibility* (Oxford, 1968), pp. 212-214.

<sup>8</sup> My use of "accountability" differs slightly from Baier's in his very useful study, "Responsibility and Action," in Myles Brand (ed.), *The Nature of Human Action* (Glenview, Illinois, 1970). See esp. pp. 103-108.

<sup>9</sup> My notion of liability responsibility is suggested by, but not equivalent to, Hart's notion of liability responsibility. See Hart, *op. cit.*, pp. 225-227.

<sup>10</sup> For instance, Mill says, "What is meant by moral responsibility? Responsibility means punishment." See *An Examination of Sir William Hamilton's Philosophy* (New York, 1874), vol. II, excerpted in Berofsky, *op. cit.*, p. 171. Cf. Schlick: "the question of who is responsible . . . is a matter only of knowing who is to be punished or rewarded," *op. cit.*, p. 61.

<sup>11</sup> Hart, *op. cit.*, pp. 227-230.

for many of the consequences of our actions, and some of these are other actions of ours. Moreover, if  $x$  were not morally responsible for hitting the child, it would be wrong to determine his blame and punishment in part according to what he *freely does*. Nor would there be any difference in respect of responsibility, blameworthiness, (moral) punishability, between  $x$  and someone who drove through a moment of *involuntary* but inconsequential recklessness. (It is very difficult to say what this difference is, but there surely is one.) Note also that it is correct to say such things as that  $x$  must bear responsibility for crippling a child and deserves the condemnation he has gotten for scarring the child's life. The most—perhaps the only—plausible interpretation of these remarks presupposes that  $x$  is morally responsible for hitting the child.

The point that " $xMRAt$ " does not entail " $xFAt$ " also holds for intentional actions. Imagine a man whose job is to guard a missile launcher. Suppose he realizes that both the duties of his office and his ordinary moral obligations require that he give his life rather than reveal the combination of the launching mechanism to unauthorized persons. Assume that as a precaution he has readily available a poison which he need only touch to commit suicide instantly. Let us also assume that his physical endurance is high. Now suppose that for two hours he is threatened with death by gunmen, severely slapped by them, and punched several times in the stomach. If, as a result, he then (intentionally) gives the combination, knowing that many thousands will probably be killed, he would surely be morally responsible, even blameworthy, for giving it. He ought to have taken the poison or held out longer. Yet it would be wrong to say that he gave the combination freely, or that his giving it was a free action. It was extracted from him only through threats and sustained force: "beaten out of him." Granted, if the beating were more severe, or went on for ten hours, the action would be even further from being free. There may also be a point at which the torture would be so severe that we could no longer consider the man morally responsible for giving the combination at the time he did, though he might still be morally responsible for not touching the poison before his assailants took hold of him and made this impossible. But where he simply fails to endure the threats and physical abuse for two hours, we

cannot reasonably say either that he freely gives the combination (or gives it of his own free will), or that he bears no moral responsibility for giving it. To be sure, if, like many philosophers, one assumes that "moral responsibility presupposes freedom," one may insist that the combination is given freely. But surely this is not plausible independently of the above doctrine: if one had no interest in protecting it, one would not say that the frightened, pained, and bleeding guard freely gave the combination.

The case of the guard is doubly interesting because it shows that even when  $x$  does not do  $A$  freely it may be true that he could have done otherwise. The guard could and in fact should have given up his life rather than reveal the combination. Another reason for denying that free actions are those one could have done otherwise than perform is that "could have done otherwise" does not admit of degrees, whereas freedom in the relevant sense does. If  $x$  joins the Army because he wants a military career and  $y$  joins only because the Army offers him a better job than anyone else, then other things being equal  $x$  joins more freely than  $y$ . It may be true that, very often, when an action is such that one could not have done otherwise, or, as I shall sometimes say, *unavoidable*, it is as far as possible from being free. But this does not imply that " $xFAt$ " is equivalent to " $x$  could have done otherwise than  $A$  at  $t$ ." It is in part because these notions have been mistakenly equated that freedom has been thought necessary for moral responsibility.<sup>12</sup>

An important question that now arises is whether " $xMRAt$ " entails that  $x$  could have done otherwise than  $A$ . We should first make explicit that "could have done otherwise" needs to be completed by the specification of two temporal variables: one designating the time of action, the other the time of avoidability, i.e., the time at which  $x$  could have done otherwise than  $A$ . Our case of the reckless driver shows that " $xMRAt$ " does not entail that at  $t$  (the time of action)  $x$  could have avoided doing  $A$ . But it does seem to entail that either at  $t$  or at some time prior to  $t$   $x$  could have done otherwise than  $A$ . This in turn raises the question of the conditions for such avoidability. For one thing, there must be (or at least must have been) some other action which  $x$  had the ability and the opportunity to perform. As in the case of the reckless driver, either there must be some alternative action(s)

<sup>12</sup> See, e.g., Campbell, *op. cit.*, especially pp. 118–119 and 133; and P. H. Nowell-Smith, "Psychoanalysis and Moral Language," reprinted in May Brodbeck (ed.), *Readings in the Philosophy of the Social Sciences* (New York, 1968), p. 712.

which, at  $t$ , he had the ability and opportunity to perform, or at some earlier time at which he could reasonably be expected to believe that he would (might) do  $A$ , there must have been some other action(s) which he had the ability and opportunity to perform and could reasonably be expected to see would (might) prevent his doing  $A$ . The chief point of this last restriction is to make clear that, when there is a question whether  $\alpha MRAt$ , it is not sufficient for "he could have done otherwise" that at some time before  $t$  there was some action which he had the ability and opportunity to perform and which would (might) *in fact* have prevented  $A$ . If this were sufficient, we would have to say that a man who is unexpectedly held up at gunpoint on a street he knows is normally safe could have done otherwise than give up his money, because he had the ability and opportunity to take a different route home. The sense in which he could have avoided giving up his money by taking another route is not the sense of "avoidability" presupposed by moral responsibility.

This is not the place to discuss ability.<sup>13</sup> What needs to be said now is that what  $\alpha$  could have done, insofar as it is relevant to his moral responsibility, is not definable simply in terms of what is physically and psychologically possible for  $\alpha$ , nor even definable in more complex non-normative terms. As the missile case suggests, where there is a question of whether  $\alpha MRAt$ , what  $\alpha$  could have done is in part determined by what a morally sound person in the situation in question might reasonably have been expected to do. There is of course disagreement over what constitutes a morally sound person. But in schematic terms, which are all our purposes require here, a morally sound person is one who holds "the right" moral principles, has "the right" moral intuitions, and, with minor exceptions, does the "right things."

Let me illustrate further how the notion of a morally sound person figures in the concept of avoidability. Suppose that kidnappers who threaten the life of his son force  $\alpha$  to give up his life's savings. If asked whether somehow he could have kept the money,  $\alpha$  could reasonably reply that he could not have done otherwise. His point would normally not be that he was physically or psychologically unable to do otherwise (nor was he, presumably), but that it would have been wrong (or unreasonable) to do so. It is also true here that

$\alpha$  does not freely give the money. This suggests that if one could not have done otherwise than  $A$ , then one did it *unfreely*. But suppose that because of a solemn promise one helps a sick person with some legal problems. Upon being thanked for one's help, one might say "I couldn't have done otherwise." This is the kind of unavoidability Kant called moral compulsion. Granted that here one implies one was not *free* to do otherwise, does one also imply that one did not do the service freely? It seems so, though one could argue that the only sense of "free" in which the action is not free is the sense in which it means "spontaneous." But we do not, at least not outside discussions of free will, speak of morally compelled actions as free in any other sense, or as done of one's own free will; we speak instead of being compelled by, say, duty, to do them. To be sure, since moral compulsion is unlike most other compulsions because, at least normally, it constitutes a justification and it is never merely an excuse, one might want to call it compulsion in a different sense. But since we do treat it as a kind of unavoidability and can easily distinguish it from other compulsions, it seems more reasonable to resist bringing in a different sense of "compulsion." We can thus seek a unified account of its various uses and avoid multiplying concepts of compulsion beyond necessity.

May we, then, say that " $\alpha MRAt$ " entails that at or at least before  $t$ ,  $\alpha$  could have done otherwise than  $A$ ? Frankfurt has ingeniously attacked this view through two examples.<sup>14</sup> First, suppose  $\alpha$  decides to do  $A$ , but is then ordered under threat to do it. If he now does it purely for the reasons he has independently of the threat, he is morally responsible for doing it, yet because of the threat could not have done otherwise. Secondly, suppose that  $\gamma$  has a way of knowing whether  $\alpha$  has decided to do  $A$  and, if he finds that  $\alpha$  has not decided to do it, can and will do whatever would justify ascribing "could not have done otherwise than  $A$ " to  $\alpha$ . Now if  $\alpha$ , who is ignorant of  $\gamma$ 's power, does  $A$  for his own reasons,  $\alpha$  would then be morally responsible for doing  $A$ , even though he could not have done otherwise.

Are these genuine counterexamples to avoidability as a necessary condition for moral responsibility? Take the first case. Suppose  $A$  is a murder. Surely if  $\alpha$  could commit it for his own reasons and without regard to the threat, then he could have

<sup>13</sup> For a very good account of this notion, see Goldman, *op. cit.*, esp. ch. Seven.

<sup>14</sup> Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy*, vol. 66 (1969), esp. pp. 832-836.

abstained from it for moral reasons and without regard to the threat. If  $x$ 's personal reasons for doing  $A$  can overshadow the threat, moral reasons for not doing it should be able to do so also. The kind of person in whom moral reasons for abstaining from  $A$  cannot outweigh personal reasons for doing it is certainly one sort we wish to—and do—hold morally responsible for doing  $A$ . Thus, surely  $x$  could have done otherwise. Frankfurt's second case is more complicated. It does show that  $x$  may be morally responsible for doing  $A$  even if, during some small time interval following  $t$  (the time  $x$  does  $A$ ),  $x$  would not have been able to do otherwise. But I hold only the restricted view that " $xMRAt$ " entails that at  $t$  (or in some cases before  $t$ )  $x$  could have done otherwise than  $A$  at  $t$ . Now suppose that, in Frankfurt's second case,  $x$  pleaded, as an excuse for  $A$ , that he could not have done otherwise. This would not do: his own decision surely did not compel him, and  $y$ 's potential compulsion was not operative. The general point here, which Frankfurt's argument overlooks, is this. If  $x$  does  $A$  at  $t$  for reasons of a kind that make him morally responsible for doing it, the fact that if he had not had those reasons and did not intend to do  $A$ , he would have been compelled to do it (whether at or after  $t$ ) for other reasons, does not entail that he could not have done otherwise at  $t$ . Frankfurt does not show that there is any possible set of reasons why  $x$  does  $A$  at  $t$  which makes it plausible to say that  $xMRAt$  and  $x$  could not have done otherwise than  $A$  at  $t$ . To be sure, in his second example it may be causally necessary, given the power of  $y$ , that  $x$  do  $A$ ; but Frankfurt does not argue, nor does he seem to assume, that this entails that  $x$  could not have done otherwise. That in fact no such entailment does hold will be supported by a number of arguments in this paper.

If Frankfurt's examples can be dealt with as I have suggested and if I am right about "could have done otherwise," then " $xMRAt$ " does entail that at (or before)  $t$ ,  $x$  could have done otherwise than  $A$  at  $t$ . The converse, however, is false. Suppose that an art student asks "Did he have to put that yellow patch there?" One might reply "No, he could have used a soft blue." In contexts like this the assessment is not moral; and if the fact that  $x$  could not have done otherwise absolves him of any kind of responsibility, it is not moral responsibility.

In most contexts, however, both questions of whether  $x$  could have done otherwise than  $A$  and

questions of whether he did  $A$  freely, are meant at least in part to ascertain whether  $xMRAt$ . My main concern is to give an account of the notion of free action embedded in these contexts and to determine its relation to the notion of moral responsibility. I shall assume that we can best understand acting freely by considering contrasts to it. For I take it that normally sentences of the form " $x$  did  $A$  freely" have the precise content they do by virtue of what, in the context, they are intended to rule out.<sup>15</sup>

## II. ACTING UNDER COMPULSION

I shall assume that the most general contrast to the notion of acting freely is that of acting under compulsion. I shall also assume that  $x$  does  $A$  under compulsion at  $t$  if and only if  $x$  is compelled to do  $A$  at  $t$  ( $xCAAt$ ), though these expressions differ in nuance. Various locutions are used in affirming that a person acted under compulsion. I have already discussed "could not have done otherwise." Consider also "I was forced to open the safe"; "He could not help using too much water, because of his obsession with cleanliness"; "She couldn't refrain from hiding her face on the bridge, because of her acrophobia"; and "He seized his dancing partner out of an irresistible impulse." Being compelled often results from being forced, coerced, pressured, or driven, but these are not mutually equivalent. Compulsions may be external or internal. The former include such things as blackmail and various threats, most notably that of death; the latter include things like addictions and certain obsessions, phobias, and unconscious desires. I shall not consider whether the sub-classes are mutually exclusive. The classification is simply for convenience.

Some of my examples raise the question whether compelled actions must be intentional. Typically, this is so. One might think that where  $x$  is compelled not to do  $A$ , e.g., not to appear in court,  $A$  may not be intentional. He might be locked up and thus physically unable to appear. It would then be false to say he intentionally fails to appear. But here his failure to appear is not an action at all: we must distinguish between an act of omission and the mere non-performance of an action. If  $x$  were blackmailed into not appearing, yet was physically and psychologically able to do so, then he intentionally abstained from appearing, though under compulsion. But where his failure to appear is due

<sup>15</sup> For a good discussion of this idea, see Cranston, *op. cit.*, esp. pp. 3-47.

solely to the physical impossibility of his appearing, this failure is surely not an action. A more problematic case would be one in which, under threat of death, a clerk hands over the money in the safe and in doing so bankrupts the firm, which, unbeknown to him, is not insured. If the clerk has no idea that he may be bankrupting the firm and thus does not do so intentionally, can we say he did so under compulsion? One might say that he bankrupts the firm only as a *result* of compulsion. A distinction between what we do under compulsion and what we do as a result of it is certainly worth drawing; but it does not follow that the clerk's bankrupting the firm is not done under compulsion. Indeed, if he were blamed by an unsympathetic stockholder for bankrupting the firm, he might reasonably protest that he couldn't help it, or even that he did so under compulsion.

One may object that since he bankrupts the firm non-intentionally he does not do so unwillingly, and that if  $x$  *CA*  $t$   $x$  must do  $A$  unwillingly. This holds for the typical cases of compulsion. But imagine a peculiar clerk who has no disposition to abstain from giving away the money in the register and no disposition to do so either. Still, if he has the normal desire to live, and, at gunpoint, gives the money to save his life, he is compelled to give it. Or, suppose a man is told under hypnosis that when a certain topic comes up during a lecture he is to attend, he will stand up and then reseal himself. If, when the topic comes up, he feels an irresistible urge to stand up and thus suddenly does so, he may or may not do so unwillingly. To be sure, the standing cannot be so automatic that it qualifies only as bodily movement. Mine is a case in which he is surprised that he stands up and immediately sits down in embarrassment. He would doubtless say that he did not know why he stood; and in some such cases he could be said both not to have stood up intentionally and to have been compelled to stand. Actions performed under compulsion, then, need be neither intentional nor performed unwillingly.

We can now work more positively toward an account of compulsion. To begin with, it seems that if  $x$  *CA*  $t$ , then no motive of personal gain is an important part of what motivates  $x$  to do  $A$ . I shall assume that, except in cases of overdetermination, in which there are two or more independent sets of

sufficient conditions, a motivational factor is an important part of what motivates  $x$  to do  $A$ , if and only if it is at least necessary for  $x$ 's doing it. If a motive of personal gain is an important part of what motivates  $x$  to do  $A$ , then surely, one wants to say, he sees himself as getting too much out of it. Perhaps  $x$  has no desire to promote  $y$ ; still, if a threat, say to expose  $x$ 's relations with the underworld, is not sufficient to make him promote  $y$  without his believing that it will yield him extra income, which he wants, then  $x$  did not make the promotion under compulsion.

One difficulty with the suggestion that motives of personal gain cannot be important parts of what motivates compelled actions is that "motive of personal gain" is vague. But the following points will indicate roughly how it is being taken. First, very often, an intrinsic want to do  $A$  is a motive of personal gain. Roughly, to say that  $x$ 's want for something is (purely) intrinsic is to say that he wants it for his own sake; i.e., his want for it is such that, if he has any reason for wanting it, no part of this reason is a belief that satisfying the want is necessary to (will, may) realize anything else that he wants, nor is any such belief a necessary condition for his retaining the want. If, e.g., threats of a drubbing would not have led the clerk to give up the money if he had not been motivated by an intrinsic want to wreak vengeance on the store's owner, he does not give it under compulsion. Motives of revenge should, for our purposes, be counted motives of personal gain. By contrast, if  $x$  has an intrinsic want to do his duty, or to preserve his life, the want is clearly not a motive of personal gain; indeed, surely no want to do what one thinks morally obligatory is a motive of personal gain. A second class of motives of personal gain consists of the typical desires for more property, power, influence, or prestige than one has. When such an increase is wanted for friends or for certain relatives, this want is also a motive of personal gain—unless  $x$  wants the increase because he thinks he is morally obligated to realize it. Nor should a desire to avoid blackmail count as a motive of personal gain, even if  $x$  has an obligation to reveal what his would-be blackmailer threatens to expose. But it is difficult to be precise about "personal gain," though at least the vast majority of such motives seem to be of the above sorts.<sup>16</sup>

<sup>16</sup> Perhaps we should require here only that a motive of personal gain not be one of  $x$ 's *main* reasons for doing  $A$ . For some uses of "compel" appear to violate my requirement. Someone might say that his family compelled him to take an extra job because they wanted more money. But it seems to me that if he would not have taken the job unless he believed he stood to gain *more* than what his family and he needed (or what he was obligated to provide for them), he is using "compel" in an extended sense.

Another important point is that if  $x$  *CA* $t$  then clearly there must be someone or something which compels him. In what I shall call the standard cases, there must be some action(s) of one or more other persons, or some condition(s) (e.g., the unavailability of food, or a psychological condition such as an irresistible desire or even a paranoid delusion), or some event(s) (such as a hail storm), which impose on  $x$  his predominant motivation for doing  $A$ . One may think that, at least in cases of compulsion by another person,  $x$  must disapprove of the compelling action(s). But this is not necessary, nor is  $x$ 's disapproving of his doing  $A$  at  $t$ :  $x$  might think it would be good for  $y$ , his too lenient employer, to threaten him with dismissal to compel him to finish an important project; and if  $y$  does so,  $x$  might not disapprove of it, or of his doing the project.

We need a further condition to distinguish compulsion both from persuasion and from mere influence. We need to capture the idea that, in the standard cases, if  $x$  is compelled he feels he "has no choice." What I suggest is that, at least by  $t$ ,  $x$  must believe that there is some state of affairs on account of which his not doing  $A$  would (might) have very bad consequences, and as a result of this belief he must also believe either (or both) (i) that his doing  $A$  is so substantially preferable to his not doing it that it would be very unreasonable for him not to do it, or (ii) that from the point of view of his own welfare, or the welfare of someone (or something) he wants to protect, his not doing  $A$  would (might) have very bad consequences. Clearly, we must put some restriction on *what* consequences are relevant. Not just any consequences  $x$  believes his not doing  $A$  would have, and considers very bad, would justify his saying he was compelled to do  $A$ . Our interest is mainly in contexts in which there is a question whether  $x$  is morally responsible for doing  $A$ , and at least in these  $x$  must believe that the very bad consequences of his not doing  $A$  are of the following sorts, or at least as serious from the point of view of human welfare, or the welfare of other sentient beings: first, death, or serious physical harm, or serious psychological harm, or crippling loss, to  $x$  or to some person(s), or social entity (such as a nation) with whose welfare he is concerned (where personal loss, e.g., financial loss, excludes

mere lack of personal gain); secondly, the destruction or serious damaging of important property; and thirdly, a serious immorality (this accommodates moral compulsion and could be omitted if we found it best to construe moral compulsion as compulsion in an extended sense).

The idea here is that, for the standard cases of compulsion, one cannot be compelled to do  $A$  unless one believes that it is in some way "too costly" not to do it. One may object that this would not account for such things as compulsive handwashing. But we must distinguish what is compulsive from what is compelled. Not all actions representing compulsive behavior are performed under compulsion; and typically the compulsive handwasher is at worst compelled to wash very frequently: not to wash *on* precisely the occasions he does. Unfortunately, it is impossible to be precise about what envisaged consequences can sustain a claim that  $x$  *CA* $t$ . But this is to be expected if we are to analyze, rather than replace, the notion of compulsion. For often the question whether  $x$  *CA* $t$  turns on whether what he believed would occur if he did not do it qualifies as "unacceptably bad" and thus as leaving him no choice. Suppose a messenger,  $x$ , gives up a briefcase which he knows contains important documents, because he is threatened by a man much bigger than he, who says he will catch  $x$  on his way home some night and make a punching bag of him. Was  $x$  compelled to give up the briefcase, or just cowardly? The main reason the answer is not clear is that there is some question whether the indefinite risk of such a beating is the kind of very bad consequence  $x$  must envisage to justify the claim that he was compelled to do  $A$ . I shall return to this problem.

One might think that if  $x$  *CA* $t$ , he must have some reason for his belief(s) to the effect that his not doing  $A$  would be too costly. But suppose that on a crowded street, and in clear view of two policemen whom  $x$  can see, a jovial, harmless-looking drunk,  $y$ , with his hands in his coat pockets, asks  $x$  for a dollar. If  $x$  thinks he is being threatened with a gun, he might give the money in trepidation and later say (to someone who asks him why he gave away his last dollar) that he was compelled to. One may say that  $x$  was merely duped. Perhaps he is duped, and certainly  $y$  does not compel him. But it

---

—or deceitfully. If one takes compulsion this broadly, then when questions of moral responsibility arise, compulsion might scarcely even count as an extenuating circumstance. Possibly we should recognize more than one sense of "compulsion"; but I am not convinced of this. Indeed, often the point of the use of "compel" just illustrated is to give the impression that one was *not* in part motivated by a motive of personal gain.



is quite reasonable to say that  $x$ 's fear of death does compel him, though one could also treat the case as one in which  $x$  merely thinks he is compelled. Indeed, he does merely think that  $y$  is compelling him; but " $x$ CA $t$ " does not entail that some *person* compels him to do  $A$ , not even a person he thinks is compelling him. (I am not attempting to analyze the special case in which a person compels  $x$  to do  $A$ ; most compulsions of this sort are coercions, though there may be coercions that are not compulsions.) Similarly, suppose  $x$  believes, but has no reason to believe, that his not doing  $A$  would be a serious immorality and does  $A$  for this reason. Is this a case of *moral* compulsion? I think not, though it would be not unreasonable to say yes, provided one kept in mind that then not all moral compulsions would be cases of unavoidability. But even if  $x$  cannot do  $A$  under *moral* compulsion unless he has reason to believe that his not doing it would be a serious immorality, we may still treat the above case as a kind of (quasi-moral) compulsion. The compelling factor is  $x$ 's conviction that he must do  $A$  to avoid committing a serious immorality; and because this is the compelling factor,  $x$  could plead extenuating, though not necessarily excusing, circumstances.

The conditions so far specified only partially capture the idea that in standard cases of compulsion  $x$  feels he has no choice. We need to add that, regarding the very bad consequences we have specified,  $x$  strongly wants to avoid them and indeed wants to avoid them substantially more than he wants to avoid doing  $A$ , or to avoid any bad consequences he believes his doing  $A$  would (might) have. Must  $x$  also regard his doing  $A$  as so substantially preferable to his not doing it that it would be unreasonable for him not to do  $A$ ? This is usually so for standard cases of compulsion, but our missile case shows that it need not be. The guard might believe the reasonable thing is to abstain from doing  $A$ , but fail to muster the motivation required to act in accordance with his better judgment.

We have still said nothing which entails that  $x$  does  $A$ , nor have we ruled out his doing it for some reason that would preclude its being compelled. I suggest we say that the wants just specified, together with the above-cited beliefs concerning the consequences of not doing  $A$ , constitute at least the main reason *why*  $x$  does  $A$ . To say that a set of  $x$ 's wants and beliefs are the *main* reason why  $x$  does  $A$  at  $t$ , is, in large part, to say that if  $x$  had been forced at  $t$  to choose between (i) acting, on the belief(s), solely to satisfy the want(s), and (ii)

acting, on the remaining belief(s) motivating his doing  $A$ , solely to satisfy the remaining want(s) motivating his doing  $A$ , he would normally have chosen (i). If, e.g., the main reason why  $x$  does  $A$  is to avoid being beaten, but part of his reason is to spare his wife anxiety, then normally, if forced to choose between doing  $A_1$ , solely because he believes it would avoid the beating, and doing  $A_2$ , solely because he believes it would spare his wife the anxiety, he would do  $A_1$ .

If the conditions so far discussed are the only ones necessary, as I believe they are for the standard cases of compulsion, then we may say that for these cases  $x$ CA $t$  if and only if, at  $t$  (and for (2)–(4) usually before  $t$  as well)

- (1) It is not the case that a motive of personal gain is an important part of what motivates  $x$  to do  $A$ ;
- (2)  $x$  believes that there is some state of affairs on account of which his not doing  $A$  would (might) have very bad consequences (where these are restricted as above);
- (3) Because of the belief specified in (2),  $x$  believes either (or both) (i) that his doing  $A$  is so substantially preferable to his not doing it that it would be very unreasonable for him not to do it, or (ii) that from the point of view of his own welfare or the welfare of someone (or something) he wants to protect his not doing  $A$  would (might) have the sorts of consequences specified in (2);
- (4)  $x$  strongly wants to avoid not doing  $A$  (or what he believes are the consequences of his not doing it), and this want is substantially stronger than any want(s) he may have to avoid doing  $A$  or to avoid any bad consequences which he believes his doing  $A$  would (might) have; and
- (5) The beliefs and wants specified in (2)–(4) constitute at least the main reason why  $x$  does  $A$ .

There are vague expressions in (1)–(5), most conspicuously "motive of personal gain" and "very bad consequences." But "compulsion" is itself vague; and as I have tried to show, often in borderline cases our hesitation in applying it results from doubt about whether  $x$  has a motive of personal gain, or about whether the consequences  $x$  believes his not doing  $A$  would have are very bad in the sense sketched above. This confirms my view that these notions figure in the concept of compulsion in the way I have specified.



There remains a more serious difficulty than the vagueness of some elements in (1)–(5). The conditions do not encompass certain compulsions we might call non-standard. These include at least cases in which *A* is not intentional or, if intentional, performed without *x*'s having the beliefs specified in (2) and (3). An example of the first sort would be *x*'s standing up during the lecture because of a posthypnotic suggestion. An example of the second kind would be *x*'s being compelled to drink ale, which he abhors, because, due to posthypnotic suggestion, he suddenly feels an overpowering thirst. Here he might drink intentionally, yet not satisfy (1)–(5). Similar possibilities are implicit in certain uses of things like drugs or implanted electrodes to control behavior, whether "directly" or through inducing sufficient motivation. I believe that such cases can be accounted for mainly in terms of the materials of the account of the standard cases of compulsion, by disjoining the following to (1)–(5):

- (6) Either (1) or, if a motive of personal gain is an important part of what motivates *x* to do *A*, it is, without his consent, induced in him by some interference with his normal functioning, such as hypnosis, implanted electrodes, drugs, or surgery;
- (7) Either (2) or there is some state of affairs which produces in *x* either simply a powerful (and possibly unconscious) desire to do *A*, or a non-motivational tendency to do *A*, i.e., one that is not a tendency to do *A* in order to realize any extrinsic or intrinsic want;
- (8) Even if, at *t*, (i) *x* should believe (as he may or may not) that his doing *A* would have a consequence of the sort specified in (2), and (ii) *x* was as strongly motivated to avoid doing *A* as would be appropriate to this belief, he would still do *A*;
- (9) Either the non-motivational tendency specified in (7), or a set of wants and beliefs of the sort cited in (2)–(4), or one or more desires of the kinds cited in (6) and (7), constitute at least the main reason why *x* does *A*.

We could simplify (8), which is meant to guarantee that the compelling force is sufficient to justify saying *x*CA*t*, by requiring that no amount of motivation possible for *x* at *t* would prevent his doing *A*. But this would be too strong. Suppose that, in our lecture case, if *x* knew his life depended on remaining seated he would have resisted stand-

ing. Does this imply that he was not compelled to stand? I think not. For if *x* had tried as hard not to stand as was appropriate in the light of what he had reason to think was at stake, e.g., damaging his reputation, we could plausibly say he was compelled to stand. But suppose that *x* was warned of the posthypnotic suggestion and that we had good reason to think it was insufficient to outweigh the appropriate degree of motivation to abstain. Here *x*'s claim to have been compelled to stand would be disallowed.

If my main points are correct, we can say that *x*CA*t* if and only if either (1)–(5) or (6)–(9), or both. The standard cases seem to be those satisfying the former set, or both sets; compulsions satisfying only (6)–(9) seem non-standard. But it might be reasonable to call some compulsions fitting only (1)–(5) non-standard and some fitting only (6)–(9) standard. My suggested distinction between standard and non-standard cases is not of major importance. It does, however, help with some puzzles.

Suppose *y* compels *x* to rob *z*'s store, which is the local heroin center, though neither *x* nor *y* has any inkling of this. Is *x* also compelled to rob the local heroin center, assuming this is a distinct action? We are pulled in both directions. No, because neither *x* nor *y* had any idea *x* was doing this. Yes, because *x* "couldn't help" robbing the local heroin center. The way out, I think, is to note that *x*'s robbing *z*'s store is a standard case of compulsion and his robbing the local heroin center is not. If we conceive both actions in terms of (1)–(5), which is natural since the paradigms of compulsion satisfy (1)–(5), we cannot account for the intuition that (because *x* couldn't help it) he was compelled to rob the local heroin center; for (1)–(5) require that *x* realize that he is (or at least might be) doing *A*. However, assuming *y*'s threat is severe enough, *x*'s robbing the local heroin center *would* satisfy (6)–(9). One can see this by noting how the two sets of conditions are related. Note first that typically, if *x*CA*t* in the sense of (1)–(5), *x*CA*t* in the sense of (6)–(9): he would obviously satisfy (6), (7), and (9); and he would typically satisfy (8) because if *x* satisfies (1)–(5) typically his motivation to avoid the consequences of *not* doing *A* would be stronger than the minimum motivation to avoid *doing A* specified by (8). Hence, typically if *x* satisfies (1)–(5) he would do *A* even if he believed it had *some* consequence of the kind specified in (8) and he were as highly motivated to abstain as (8) requires. Secondly, for the same sorts of reasons, if *x*CA*t* in the sense of (1)–(5), and either *by* or *in* doing *A* he

must, at least in fact, do  $A_1$ , then in most cases he is also compelled to do  $A_1$ , at least in the sense of (6)–(9). For instance, if  $x\text{CA}t$  in the sense of (1)–(5) then normally his motivational tendency to do  $A$  will create a (possibly non-motivational) tendency to do whatever he must do in (or by) doing  $A$ , and the latter tendency will normally have sufficient strength to make  $x$  satisfy (8). It is the in-relation that holds between  $x$ 's robbing  $x$ 's store and  $x$ 's robbing the local heroin center. But suppose  $x$  is compelled to fire a gun and by doing so he creates a smell of sulphur. Is he also compelled to create a smell of sulphur? Again we may be pulled in two directions; for  $x$ 's creating a smell of sulphur may not be intentional, and typically the actions we speak of as compelled are. Nevertheless, actions that satisfy (6)–(9) are surely compelled, and this accounts for our tendency to regard as compelled most actions that are performed in (or by) performing one that is compelled in the sense of (1)–(5). What I am saying implies that when  $x\text{CA}t$  he is also compelled to do indefinitely many other things. But typically few if any of these concomitant actions are worth mentioning (which largely accounts for the peculiarity of saying, of such actions, that  $x$  is compelled to do them); and though they must not be overlooked, they pose no special problem here.

In closing this section I want to make a general point about compulsion. Where we take  $x$ 's doing  $A$  to be morally wrong, the more seriously wrong we take it to be, the more stringently we tend to interpret the conditions; e.g., the more evidence we require to be convinced that  $x\text{CA}t$ . This is not surprising, since we allow compulsion to count as an extenuating circumstance even where it does not absolve  $x$  of moral responsibility. There is, then, a normative element in many important uses of "compel." The above account of compulsion—and probably any realistic account of it, or of free action construed as its opposite—should be interpreted in the light of this, despite the complexity it adds. If  $x$ 's doing  $A$  is a mere peccadillo, it would be unreasonable to interpret the conditions stringently, disallowing, say, that the psychological drubbing he believed he would get from his wife was "unacceptably" bad; and if  $x$ 's doing  $A$  is a grave wrong, such as a murder of a competitor, it would be unreasonable to allow  $x$ 's anticipated loss

of a good client to count as a very bad consequence of the relevant kind.

### III. ACTING FREELY AND ITS RELATION TO MORAL RESPONSIBILITY

If it is true that the most general contrast to "acting freely" is "acting under compulsion" and that to say that  $x$  did  $A$  freely is to rule out some kind of compulsion indicated in the context, then perhaps we can understand free action as (intentional) action that is not compelled.<sup>17</sup> One difficulty is that even where  $A$  is intentional "free" and "not compelled" may seem to be contraries, not contradictories. When one (intentionally) turns the page of a book one is reading, must one do this either freely or under compulsion? To say one does it freely would suggest the possibility of compulsions whose presence here is ordinarily out of the question. This is the sort of consideration that led Austin to say "No modification without aberration."<sup>18</sup> But if, where  $A$  is intentional, " $x\text{FA}t$ " is equivalent to "It is not the case that  $x\text{CA}t$ ," it would seem that we must say one freely turned the page. For such actions are both intentional and un-compelled.

The simplest way to deal with this objection is to distinguish between what is true and what there is point in saying. It is true that this paper is white, but there would ordinarily be no point in saying this and it would thus be odd to say it in an ordinary situation. So with "free." A more cautious reply would be this. *Given* that it is appropriate to ask whether  $x$  did  $A$  freely—and I take this question to presuppose that  $A$  was intentional—he did it freely if and only if he was not compelled to do it. The plausibility of this suggestion is confirmed by the fact that questions of whether  $x$  was compelled to do  $A$  are appropriate and inappropriate in roughly the same cases in which "Did  $x$  do  $A$  freely?" is appropriate and inappropriate. Indeed, normally "I was not compelled to turn this page" would have the same sort of oddity as "I freely turned this page," and for the same reason: that in the context there is no hint of what compulsion is being ruled out and hence no point in making the statement. Apparently, then, we can meet the above objection to construing free actions as un-compelled intentional actions.

<sup>17</sup> Schlick held that "Freedom means the opposite of compulsion" (Berofsky, *op. cit.*, p. 59). Hume and others have held a very similar view. What I have tried to do is, in part, to fill the gap other compatibilists have left in our understanding of compulsion.

<sup>18</sup> J. L. Austin, "A Plea for Excuses," reprinted in Herbert Morris (ed.), *Freedom and Responsibility* (Stanford, 1961), p. 12.

Another objection to this proposed equivalence is that whereas there are degrees of freedom with which one can do *A*, there are no degrees of compulsion. We can meet this objection as follows. Perhaps "degrees of compulsion" does not have a common use; but since there are degrees of compelling force, i.e., the force used to compel *x*, we can give "degrees of compulsion" a clear use, and indeed it is a merit of the above account of compulsion that it enables us to make clear sense of this comparative notion. Thus, we can suppose both that *x* freely does *A* when it is intentional and uncompelled, and that he does it more freely in inverse proportion to the number and intensity of the elements present which, when they are all present in the appropriate degree, render the action compelled. We can also say that when  $xCA_t$ , the greater the degree of force used beyond the minimum required for the compulsion, the further *x* is from doing *A* with any degree of freedom and the greater the degree of the compulsion. Thus, for the standard cases of compulsion, other things being equal: (a) the more *x* intrinsically wants to do *A* or believes he will gain in some other way from doing it, the lower the degree of his compulsion (the less unfree he is) in doing it; (b) the less his motivation to do *A* is "pathologic" or externally imposed, the lower the degree of his compulsion in doing *A*; (c) the worse he believes the consequences of his not doing *A* to be, the *higher* the degree of his compulsion (the less free he is) in doing *A*; (d) the more he wants to avoid these consequences, the higher the degree of his compulsion in doing *A*; and (e) the worse he believes his least objectionable alternative, the higher the degree of his compulsion in doing *A*. For non-standard cases of compulsion, other things being equal, the greater the motivation that would be required for *x* to resist doing *A*, or the further his resisting it is from the strongest motivation to resist he is capable of, the higher the degree of his compulsion in doing *A*.

Can the account of compulsion also help to explain the relation between free action and moral responsibility? I think it can. Recall that *x* can be morally responsible for doing *A* even if he did not do it freely, but not if, in addition, it was unavoidable.

Our missile guard acts under compulsion, but the action is not unavoidable, since he could have done otherwise even at the time he gave the combination. Admittedly, it is slightly odd to say that though  $xCA_t$  he could have done otherwise; but I

believe this is because the vast majority—and the paradigms—of compulsion are also cases of unavoidable. But when is an action unavoidable? And under what conditions does an action that is compelled—and thus also unfree—but not unavoidable, absolve one of moral responsibility?

Part of the answer to the first question has already been foreshadowed. We normally do not (and we surely should not) accept a claim that *x* could not have done otherwise if we believe a morally sound person in the relevant circumstances could reasonably have been expected to do otherwise. Taking this point together with the account of compulsion, we have the materials for an approximate non-hypothetical account of "could not have done otherwise" in the sense relevant to moral responsibility. Roughly, *x* could not have done otherwise than *A* at *t* if and only if  $xCA_t$  and it is not the case either (i) that *x* could reasonably have been expected to avoid (or try to avoid) the situation in which *A* occurred, or (ii) that if a morally sound person had been in the relevant situation he could reasonably have been expected to do otherwise. To be sure, if a psychotic uses up the water during a shortage, because of his obsession with cleanliness, we might grant that this action was unavoidable. But such cases do not violate (ii), since surely a morally sound person could not have been in this situation: here a purely personal desire outweighs a clear moral obligation. There is no simple way, of course, to unpack the notion of what a morally sound person could reasonably be expected to do; but here I simply want to show what would be required for a detailed analysis of unavoidability.

Our question about the relation between moral responsibility and free action is more difficult. Part of the answer is implicit in the discussion of unavoidability: this sort of compulsion is sufficient to absolve *x* of moral responsibility. But where an action is compelled and hence unfree, yet not unavoidable, how do we (and should we) decide whether the compulsion is sufficient to absolve *x* of moral responsibility?

One consideration has already been suggested by our reflections: the moral gravity of the action in the situation. Suppose that, because he is beaten and threatened with the destruction of his store, *x* reveals the whereabouts of a person he knows his threateners will then murder. His doing so under compulsion would not absolve him of moral responsibility for his action. But if, under the same threat, he gives them the \$75 in the cash register,

then normally he would not be morally responsible. Clearly in the first case the action is of greater moral gravity than in the second, where the moral gravity of an action, *A*, in a given situation, is primarily a function of (a) the degree of harm to human beings which a reasonable person would, in the relevant circumstances, believe it would (or might) do, and (b) the strength of the obligation he would normally have (apart from the compulsion) to abstain from *A* in the situation in question; (b) in turn is a function of such things as the number and kinds of rights *A* violates.

These remarks suggest a second consideration: the degree of compelling force that brings *A* about. Suppose *x* is threatened with death by two gunmen if he does not immediately hand them the purse of a lady he is escorting. Normally, he would not be morally responsible for doing this. But if a twelve-year-old boy with a small pocket knife made the same threat, then normally an able-bodied man who made no attempt to resist would be morally responsible for handing over the purse, even if he were ultimately excused. I have suggested some ways of measuring degree of compelling force, though in practice this is difficult, especially for internal compulsions. Regarding these, I shall say only that we can get an idea of the degree of compelling force by combining our knowledge of some of *x*'s behavior with general considerations about his psychological (or physical) make-up. With external compulsions we can judge partly on the basis of how much harm *x* believes his not doing *A* will (may) cause him or others; how much pain or fear he withstands before doing *A*; and the degree to which he attempts to escape performing the action: such attempts not only suggest that more force was used than would have been had *x* done *A* at the first opportunity; they also have independent extenuating import because they show that *x* tried to avoid doing *A*. There are other factors as well, though none of them admits of precise measurement. Perhaps the best we can hope for is a way of rationally deciding, for some cases of compulsion, whether there was more force used than in some other cases. But even this is significant in assigning moral responsibility, particularly since we can make sense of interpersonal as well as intrapersonal comparisons.

A third consideration concerns how much reason *x* has for the beliefs and wants motivating the compelled action. Suppose *x* is occasionally paranoid and believes, without reason, that if he does not dismiss an employee under him (whom *x*'s

boss dislikes) *x*'s boss will dismiss *x*. In some such cases *x* might be compelled, by his own delusion, to dismiss the person. But other things being equal, the less reason *x* has for the motivating beliefs and wants (the wants and beliefs of the forms specified in (2)–(4)), the less his being compelled to do *A* extenuates his moral responsibility for it. One may object that if *x*—and his motivating beliefs and wants—are psychotically irrational, then far from being more responsible, *x* is so disturbed that he bears no moral responsibility for *A*. But when this is so it will usually turn out that *x* could not have done otherwise, and then we are no longer talking about an action which is compelled but not unavoidable. But even supposing *x* could have done otherwise, it does not follow that *x*MR*A**t*. As I have shown, avoidability is only a necessary, not a sufficient, condition for moral responsibility.

Sometimes utilitarian considerations also do and should enter into determining when a person is morally responsible for an action which is compelled yet not unavoidable. This applies more to legal than to moral responsibility. But sometimes it is relevant to ask what would be the effect, on the welfare of the victim and others, of considering him morally responsible in such a case, and what would be the effect on the welfare of people in general of adopting a policy of considering persons morally responsible under the conditions in question. Suppose that *y*, who is a good deal bigger than *x*, threatens *x* with a beating if *x* does not give *y* *z*'s car, where *x*, who has borrowed the car, knows *z* will be seriously inconvenienced by this, even though *z* has the car insured. We might well allow that *x* was compelled to give up *z*'s car, yet be uncertain whether to consider *x* morally responsible for doing so. Granted that the sorts of utilitarian considerations I have mentioned are relevant to deciding whether to hold *x* morally responsible, e.g., whether to demand that he defend his action, are they relevant to deciding whether he is morally responsible? I believe they may be. For people usually do not like to be considered morally responsible (prima-facie blameworthy), even if they believe they will be excused; and if we adopt a policy of considering them morally responsible in cases like the above, this may well increase their tendency to resort to violence to avoid the actions in question. That might be undesirable on balance and might warrant us in deciding that, in such borderline cases, *x* is not morally responsible. But whether utilitarian considerations are ever the reason we consider *x* morally responsible in non-

borderline cases is a large question, and I shall not try to answer it here.

#### IV. AVOIDABILITY, COMPATIBILISM, AND DISCLAIMERS OF MORAL RESPONSIBILITY

Let me now draw out some implications of my main points. I have argued that a person's being morally responsible for an action does not entail that he performed it freely, though it does entail that there was at least some time at which he could have done otherwise. One reason this negative thesis is important is that philosophers have commonly assumed, and to my knowledge not seriously questioned, that in clarifying the idea of acting freely we are clarifying a necessary condition for moral responsibility. I do not deny that often  $x$ 's doing  $A$  under compulsion, and therefore unfreely, absolves him of moral responsibility for it; and I have tried to sketch an account of the conditions under which compulsion does absolve. But only an action's being unavoidable—and therefore unfree in a very strong sense—automatically absolves  $x$  of moral responsibility. One may protest that I have merely shown that many philosophers do not use "free" in the ordinary sense. I doubt whether many of them use it as a technical term in the contexts that chiefly concern us, those in which there is a question of someone's moral responsibility for an action. But if "free" is being used as a technical term in important contexts, and this is not generally recognized, there is a danger of confusion. Moreover, if, in actual moral discourse, we distinguish the concepts of free action, unavoidable action, and action for which the agent bears moral responsibility, then conflating these, as a good many philosophers have, not only invites confusion but reduces the descriptive power of our system of moral concepts. But let me indicate more important consequences of my conclusions.

The first is that the notion of avoidability is in part a moral notion. It is not, e.g., definable simply in terms of physical and psychological ability, or absence of causal determination. For at least very often, to ascertain whether  $x$  could have done otherwise, we must determine, or assume we know, whether a morally sound person could reasonably be expected to do otherwise in the relevant situation. Moreover, where considerations of the freedom of an action bear on whether  $x$  was morally responsible for it, here too certain moral notions

are inescapable. We must consider, e.g., the moral gravity of his action in the situation. Some philosophers, especially those who have equated " $xMRAt$ " and " $xFAAt$ ," have hoped to explicate in non-moral terms at least the notion of avoidability,<sup>19</sup> which is generally taken to be necessary and sometimes also sufficient for moral responsibility. But neither this notion, nor the conditions under which acting unfreely absolves one of moral responsibility, can be explicated without the burden of employing moral terms, nor justifiably applied in practice without the onus of making, or having to assume, certain moral judgments. It is true that an action's being compelled is always an extenuating circumstance. Yet even the degree of extenuation is in part a function of moral considerations.

A second general consequence of my position can be brought out by contrasting it with the incompatibilist claim that  $xFAAt$  only if there is no set of laws of nature which, together with certain antecedent condition statements true of  $x$  at least by  $t$ , entails that  $x$  does  $A$  at  $t$ . On my view, " $FAAt$ " is compatible with, though it does not entail, the existence of such a set of laws and antecedent condition statements. This gives compatibilism a clear advantage. For on an incompatibilist view, to know, or even to have good reason to believe, that  $xFAAt$ , one must know, or at least have good reason to believe, the sweeping negative existential claim that there is no set of laws and true antecedent condition statements of the specified kind. But surely we are never in a position to know, or even have good reason to believe, such a claim about our actions.

Consider also the positive account provided by incompatibilism. On one kind of incompatibilism, to say that  $xFAAt$  is simply, or primarily, to rule out  $A$ 's being determined; whereas on my account free action can be understood in contrast to the diverse compulsions encompassed, and unified, by the necessary and sufficient conditions for " $xCAAt$ ." On a more sophisticated incompatibilist account, " $xFAAt$ " might be understood as ruling out " $xCAAt$ ," but cases in which  $x$ 's doing  $A$  is entailed by a set of laws and true antecedent conditions statements would be treated as cases of compulsion, or at least of unavoidability. But if I am even roughly correct about the notion of compulsion, to treat all such cases thus would be wrong. Moreover, there are surely some who understand the notions of compulsion and unavoidability without even being

<sup>19</sup> See, e.g., Robert J. Richman, "Responsibility and the Causation of Actions," *American Philosophical Quarterly*, vol. 6 (1969), p. 195.

aware that actions might be determined, since they lack the relevant concept of nomic determination; and of those who are aware that actions might be determined, certainly no one other than an incompatibilist (or even an incompatibilist in his everyday affairs) would automatically regard such cases as instances of compulsion. Whether an action is compelled and whether it is avoidable both depend, as I have tried to show, on the *kind* of thing that "brings it about," not on *whether* the action is brought about according to "deterministic" laws.<sup>20</sup>

A third consequence of my position has some general significance beyond its purely philosophical implications. If "*xCAI*" is compatible with "*xMRAI*," then we may reassess the impact of recent psychological discoveries—and claims—suggesting that a great many of our actions, far more than we usually suppose, are compelled and thus unfree. Broadly speaking, I refer to "neurotic" and apparently normal behavior, not to "psychotic" behavior, though no doubt many exaggerate the extent to which anything that can plausibly be called mental illness renders the actions said to result from it compelled. This century has seen a tendency on the part of the educated to regard fewer and fewer of our everyday actions

as free, and to attribute many of the allegedly vitiating compulsions to irrational (often unconscious) psychological forces, or to pervasive social pressures. The tendency has probably been strengthened by a widespread incompatibilist construal of the progress, or purported progress, of psychology in discovering the springs of human action. We may certainly dispute whether this tendency to diminish the domain of free action is justified. In the light of my account of compulsion, I suspect that the extent to which the actions of normal persons are compelled by psychological forces or social conditions is often exaggerated, though my account makes clear why it is often natural to make unwarranted extensions of the notion of compulsion, and corresponding unwarranted encroachments upon the notion of freedom.\* But however that may be, compulsion simpliciter does not defeat moral responsibility; and it is a moral question, not a psychological question, under what conditions an action's being compelled and hence unfree absolves the agent of moral responsibility for it. Insofar as one deplores the widespread tendency to try to avoid moral responsibility by vague appeals to compulsion, or facile claims of impaired freedom of action, this should be a welcome conclusion.<sup>21</sup>

*The University of Nebraska, Lincoln*

*Received December 11, 1972*

<sup>20</sup> It is worth noting that one reason many philosophers have resisted compatibilism may be that it has been defended mainly by writers in the empiricist tradition, most of whom have been utilitarians of some kind. But compatibilism does not entail any particular moral theory; hence objections to utilitarianism are not necessarily objections to compatibilism.

<sup>21</sup> For helpful comments on an earlier version of this paper, I want to thank Bernard Berofsky, John Exdell, David Gerber, Norman C. Gillespie, Hardy E. Jones, Martin Perlmutter, and H. Van R. Wilson.

## II. EPISTEMIC DEFEASIBILITY

MARSHALL SWAIN

IN the recent literature of both ethics and epistemology the concept of defeasibility has played an important role. In ethical contexts, for example, it is hardly possible to make sense of the distinction between *prima facie* and absolute obligation without a clear notion of defeasibility.<sup>1</sup> Similarly, in epistemological contexts a distinction between *prima facie* and absolute justification has been used to advantage by a number of authors.<sup>2</sup> Unfortunately, the conditions under which a given obligation or justification is defeasible have not been specified in a way that meets with unanimous approval.

One of the more interesting applications of the concept of epistemic defeasibility concerns the perplexing problem of analyzing the concept of knowledge. It is well known that it is not sufficient for a man's having knowledge that he have justified true belief; it is also well known that the problem of specifying what else is necessary is exceedingly difficult. One suggestion concerning the "fourth" condition of knowledge which has received a good deal of attention is that for a man to have knowledge his justification must be indefeasible.<sup>3</sup> Even though no adequate characterization of indefeasibility has been provided, I am now convinced that this approach to the problem of knowledge is the correct one. In any event, I shall assume here that the analysis of knowledge in terms of indefeasible justification is the one to be adopted. Granted this assumption, we can specify a partial criterion for

an adequate explication of the concept of defeasibility:

- (C) An explication of defeasible justification is adequate only if for any person *S*, and any proposition *h*, if (1) *h* is true and (2) *S* is justified in believing *h* and (3) *S* believes that *h* on the basis of his justification, then *S* knows that *h* if and only if his justification for *h* is indefeasible in accordance with that explication.

Given (C) we can "test" a proposed analysis of defeasibility by checking various agreed upon cases of knowledge and nonknowledge. A natural protest to this procedure is that to "test" an explication of defeasibility we shall need some independent way of deciding whether a man has knowledge in addition to merely having justified true belief, and we have no way of doing this. In general I would agree that it is a poor criterion indeed that directs us to test a proposed explication of one murky concept in terms of an equally difficult one. But, in this particular instance, I believe that the criterion (C) can be useful. There is a remarkable amount of agreement among philosophers and others concerning those kinds of situations that are to be called instances of knowledge and those that are not. With appropriate reservations, I see no great harm in relying upon such widespread intuitions with respect to knowledge in order to generate an analysis of defeasibility.

<sup>1</sup> A classic use of this distinction occurs in W. D. Ross, *The Right and the Good* (Oxford, 1930). For a precise treatment of the distinction and further references see Roderick Chisholm, "The Ethics of Requirement," *American Philosophical Quarterly*, vol. 1 (1964), pp. 147-153.

<sup>2</sup> One recent example is John Pollock, "Perceptual Knowledge," *The Philosophical Review*, vol. 80 (1971), pp. 287-319.

<sup>3</sup> Each of the following is an attempt to construct what I am calling a "defeasibility" analysis of knowledge, even though not all of the authors of these papers explicitly characterize their views in terms of defeasibility: Ernest Sosa, "The Analysis of 'Knowledge that *p*,'" *Analysis*, vol. 25 (1964), pp. 1-8; Keith Lehrer, "Knowledge, Truth, and Evidence," *Analysis*, vol. 25 (1965), pp. 168-175; Keith Lehrer and Thomas Paxson, "Knowledge: Undefeated Justified True Belief," *The Journal of Philosophy*, vol. 66 (1969), pp. 225-237; Ernest Sosa, "Propositional Knowledge," *Philosophical Studies*, vol. 20 (1969), pp. 33-43 and "Two Conceptions of Knowledge," *The Journal of Philosophy*, vol. 67 (1970), pp. 59-66; Gilbert Harman, "Induction" in Marshall Swain (ed.) *Induction, Acceptance, and Rational Belief* (Dordrecht, 1970), pp. 83-100; Keith Lehrer, "The Fourth Condition of Knowledge: A Defense," *The Review of Metaphysics*, vol. 23 (1970), pp. 122-128; Peter D. Klein, "A Proposed Definition of Propositional Knowledge," *The Journal of Philosophy*, vol. 68, pp. 471-482; Marshall Swain, "Knowledge, Causality, and Justification," *The Journal of Philosophy*, vol. 69 (1972), pp. 291-300, and "An Alternative Analysis of Knowledge," *Synthese* vol. 23 (1972), pp. 423-442; Risto Hilpinen, "Knowledge and Justification," *Ajatus* vol. 33 (1971), pp. 7-39.

In what follows I shall consider, and reject, a number of proposed explications of the concept of defeasible justification. I shall then present what I think is an acceptable explication of that concept. I shall rely upon the criterion (C) as an aid in judging the adequacy of various proposals.<sup>4</sup>

## I

Let us begin by considering a proposal made by Roderick Chisholm in his paper "The Ethics of Requirement."<sup>5</sup> Chisholm provides us with an explication of the ethical concept of defeasibility, but suggests that an analogous explication can be given for the epistemic notion of defeasible justification. Taking the expression "*p* requires *q*" as primitive, and letting '*p*', '*q*', and '*s*' stand for events or states of affairs, Chisholm first defines the ethical notion of overriding:

- (D1) There is a requirement for the state of affairs *q* which has been overridden = df. There are states of affairs *p* and *s* such that (i) *p* occurs and *p* requires *q* and (ii) *s* occurs and the joint occurrence of *p* and *s* does not require *q*.

Then, Chisholm gives us the following definition of ethical defeasibility:

- (D2) A requirement for the state of affairs *q* is defeasible = df. There is a state of affairs *p* such that *p* requires *q* and this requirement may be overridden.

To obtain the epistemic analog of these definitions, we can take as primitive the expression "*e* justifies *h*," where '*e*' and '*h*' stand for propositions. Then we can say:

- (D3) There is a justification for *h* which has been overridden = df. There is a body of evidence *e* and a body of evidence *e'* such that (i) *e* is true and *e* justifies *h* and (ii) *e'* is true and the conjunction of *e* and *e'* does not justify *h*.

We can then say that a justification for *h* is defeasible as follows:

- (D4) A justification for *h* is defeasible = df. There is a body of evidence *e* such that *e*

is true and *e* justifies *h* and this justification may be overridden.

In opposition to those justifications that are defeasible are justifications that are indefeasible. It seems natural, given (D4), to define this concept in the following way:

- (D5) A justification for *h* is indefeasible = df. There is a body of evidence *e* such that *e* is true and *e* justifies *h* and this justification cannot be overridden.

But the notorious word "cannot" presents us (as usual) with a problem. Presumably, to say that a justification for *h* cannot be overridden is just to say that there cannot be any body of evidence *e'* such that *e'* in conjunction with the justifying body of evidence *e* fails to justify *h*. But how are we to understand the requirement that there cannot be any such body of evidence *e'*? We might take this to mean that it is not logically possible that there is such a body of evidence. However, criterion (C) shows us that this conception of indefeasibility is not satisfactory. The analysis of knowledge associated with this conception would be the following:

- (D6) *S* knows that *h* iff (i) *h* is true, (ii) *S* is justified in believing that *h* (that is, there is a true body of evidence *e* such that *S* is justified in believing *e* and *e* justifies *h*), (iii) *S* believes that *h* on the basis of his justification and (iv) *S*'s justification for *h* is indefeasible (that is, it is not logically possible that there is a body of evidence *e'* such that the conjunction of *e* and *e'* fails to justify *h*).

Rather obviously, (D6) is too strong. One way of seeing this is as follows. If *h* is a contingently true proposition, then it is logically possible that *h* is false. But if it is logically possible that *h* is false, then it is logically possible that there is some body of evidence *e'* such that the conjunction of *e* and *e'* fails to justify *h*. For, the denial of *h* would constitute such evidence—it is clear that the conjunction of *e* with not-*h* would fail to justify *h*. So, (D6) will have the result that we can only have knowledge of noncontingent propositions. This is not a happy result.

Another way in which we might interpret

<sup>4</sup> In this criterion, and in the remainder of the paper, the question of what constitutes a man's justification is left open to interpretation. In some of the proposals to be considered (specifically, those in Sect. II) a man's justification for a proposition *h* is limited to some body of evidence *e* that he possesses. In the remaining proposals, there is allowance for the possibility that a man may have a justification for *h* even though his justification does not consist of some body of evidence.

<sup>5</sup> Chisholm, *op. cit.*, especially pp. 147–149.



"cannot" in (D5) is to say that it is not physically possible that there should be a defeating body of counter-evidence  $e'$ . We could then revise our definition of knowledge (D6) by replacing (iv) with

- (iva)  $S$ 's justification for  $h$  is indefeasible (that is, it is not physically possible that there is a body of evidence  $e'$  such that the conjunction of  $e$  and  $e'$  fails to justify  $h$ ).

However, this analysis of knowing suffers from defects similar to the previous one. Suppose  $h$  is the proposition that there is a red ball in front of me. It is surely not physically impossible that  $h$  is false—the conjunction of the denial of  $h$  with the Laws of Nature would not yield a contradiction. Hence, any justification I might have for believing that  $h$  is true would be defeasible, and I could never know of any red ball (nor of anything else) that it is in front of me. The sceptic would enjoy this conclusion, but I do not.

Yet another way in which we might interpret "cannot" in (D5) is to say that there is *in fact* no true body of defeating counterevidence  $e'$ . That is, we can say that a man's justification for  $h$  is indefeasible just in case there is in fact nothing about the world that overrides that justification. This is compatible with saying that it is physically possible that there should be such counterevidence and that it is logically possible that there should be. To incorporate this suggestion, we can replace (iv) with

- (ivb)  $S$ 's justification for  $h$  is indefeasible (that is, there is no true body of evidence  $e'$  such that the conjunction of  $e$  and  $e'$  fails to justify  $h$ ).

The resulting definition of knowledge has a great deal to recommend it. For one thing, this definition is immune to the recalcitrant problems raised by Edmund Gettier in his paper "Is Justified True Belief Knowledge?"<sup>6</sup> Gettier provides examples that prove the "traditional" analysis of knowing as justified true belief to be too weak. It is readily established that in each of his examples, and in any situation having the structure of his examples, the replacement of condition (iv) by (ivb) in the definition (D6) renders it strong enough to avoid

the problem. For, in each of his cases there is some true body of evidence  $e'$  such that the conjunction of  $e'$  with  $S$ 's evidence  $e$  fails to justify  $h$ .

Despite this meritorious feature, an analysis of knowing incorporating (ivb) will not suffice, for it is too strong. Hence, criterion (C) will lead us to reject this proposal. The following example will make this clear. Suppose that  $S$  attends a wedding ceremony in which two friends of his become married. The ceremony is performed by the Bishop and is performed without any errors. It would certainly seem that  $S$  knows, after having witnessed the ceremony, that his friends are married. But we can easily imagine the world being such that  $S$ 's justification is defeasible in accordance with (ivb). Imagine, for example, that at the time the ceremony is performed, but unknown to anyone involved in the ceremony (including  $S$ ), the Cardinal goes insane. He has long harbored a suppressed hatred of the Bishop, and in his insanity falsely denounces the Bishop as a fraud who is not authorized to marry anyone. There will then be a true body of counterevidence  $e'$  (namely, the statement "The Cardinal says that the Bishop is a fraud") such that  $e'$  in conjunction with  $S$ 's evidence  $e$  would fail to justify the proposition that the people are married. This suffices to show that the analysis of knowing incorporating (ivb) is too strong. Since (1), (2), and (3) of criterion (C) are clearly satisfied here, we should reject (ivb) as an explication of indefeasibility.

One thing we may notice about this example is that the presence in the world of the defeating counterevidence  $e'$  is an entirely unexpected fact. That is,  $S$ 's evidence  $e$  does not seem to give support one way or the other to the proposition that the Cardinal is, or is not, saying insanely that the Bishop is a fraud. If this is too strong, we can at least say that  $S$ 's evidence  $e$  does not *justify* either  $e'$  or its denial. Perhaps, then, we can weaken the defeasibility condition (ivb) by restricting the range of admissible defeating counterevidence to those true propositions for which the evidence  $e$  is strongly negative. If we revise the definition by incorporating this restriction the resulting analysis is essentially the same as one considered by Keith Lehrer and Thomas Paxson.<sup>7</sup> The revision can be effected by exchanging (ivb) for:

<sup>6</sup> Edmund Gettier, "Is Justified True Belief Knowledge?," *Analysis*, vol. 23 (1963), pp. 121–123. Reprinted in A. Phillips Griffith (ed.), *Knowledge and Belief* (Oxford, 1967), and in Roth and Galis (eds.), *Knowing: Essays in the Analysis of Knowledge* (New York, 1970).

<sup>7</sup> Lehrer and Paxson, *op. cit.*, especially pp. 227–228.

- (ivc) *S*'s justification for *h* is indefeasible (that is, there is no true body of evidence *e'* such that (a) *e* justifies *S* in believing that *e'* is false, and (b) the conjunction of *e* and *e'* fails to justify *h*).

Unfortunately, this straightforward way of incorporating the restriction does not work. Lehrer and Paxson show that (ivc) reduces to (ivb).<sup>8</sup> Suppose there is some statement *r* such that *e* completely justifies *S* in believing *r* to be false, and in believing *h* to be true, but *r* is utterly irrelevant to the justification of *h* by *e*. Then, let *q* be any true statement whatever such that the conjunction of *e* and *q* fails to justify *h* but such that *S* is *not* justified in believing *q* to be false on the basis of *e*. Then, consider the conjunction of *r* and *q*. Since *S* is completely justified in believing *r* to be false, he is completely justified in believing the conjunction of *r* and *q* to be false. But, since *r* is utterly irrelevant to the justification of *h* by *e*, the conjunction of *r* and *q* and *e* fails to justify *h*. So, there will be a true body of evidence *e'* (namely, *q* & *r*) such that *e* justifies *S* in believing *e'* to be false but such that the conjunction of *e* and *e'* fails to justify *h*. This will hold, moreover, for *any* *q* that satisfies the conditions on *e'* in (ivb). We can handle this kind of problem by revising (ivc):

- (ivd) *S*'s justification for *h* is indefeasible (that is, there is no true body of evidence *e'* such that (a) *e* justifies *S* in believing *e'* to be false and (b) the conjunction of *e* and *e'* fails to justify *h* and (c) if *q* is a logical consequence of *e'* such that the conjunction of *q* and *e* fails to justify *h* then *e* justifies *S* in believing *q* to be false.)<sup>9</sup>

In the above, *q* is a logical consequence of (*q* & *r*) but *S* is not justified in believing *q* to be false. Hence the conjunction of *r* and *q* will not qualify as an instance of defeating counterevidence *e'*. This seems an appropriate result.

The analysis of knowledge that incorporates (ivd) is a promising one, since it avoids not only the Gettier examples but also the wedding example above. Hence, in accordance with criterion (C), we have some reason to think that (ivd) can serve as an explication of defeasibility. Unfortunately, this analysis of knowing is both too weak and too strong.

<sup>8</sup> *Ibid.*, pp. 228–229.

<sup>9</sup> This is the definition of defeasibility proposed by Lehrer and Paxson, *ibid.*, p. 231.

<sup>10</sup> This example, and the following one, are similar to some examples given by Sosa in "Two Conceptions of Knowledge," *op. cit.*, pp. 60–62.

To see that this analysis is too strong, we can consider the following: Suppose that *S* has just thrown a rock at a window. It is a large rock, an ordinary window, and *S* can see that the rock is going to hit the window. Moreover, the rock is going to hit the window, and when it does the window will break as a result. It seems clear that *S* knows the window will break. For one thing, there does not appear to be any true defeating counterevidence available. We can, however, imagine some additional bizarre circumstances such that there would be defeating counterevidence in accordance with (ivd). We can note, first, that since *S* is justified by his evidence in believing that the window will break, he is also justified in believing that he will see the window break and hear it break. Now, suppose that *S* has a peculiar nervous disease hitherto unknown to the human race. This disease has never before manifested any symptoms in *S*, but he is about to be afflicted with some symptoms, namely, total visual and auditory paralysis. Let us also suppose that *S* will be afflicted by these symptoms at the instant the window breaks. So, *S* will neither see nor hear the window break—let us call this *e'*. It seems clear that *e'* in conjunction with *S*'s evidence *e* fails to justify belief that the window will break, *S* is justified in believing *e'* to be false, and we may suppose that every logical consequence of *e'* satisfies condition (c) of (ivd). Thus, we should conclude that *S* does not know that the window will break, contrary to the fact that he does know that it will break. The fact that he will be afflicted with sensory paralysis has no effect whatever on his knowing this.<sup>10</sup>

The following example structure suffices to show that the analysis of knowing incorporating (ivd) is also too weak. Suppose that *S* has some body of evidence *e* such that *e* justifies *h*. But suppose that even though *h* is true and *S* believes that *h* he does not know *h* because there is some true counterevidence *e'* such that *S* is justified in believing *e'* to be false and *e'* overrides his justification. Then, suppose there is *also* some true counterevidence *e''* available such that *e''* in conjunction with *e* fails to justify *h* but *S* is *not* justified in believing *e''* to be false. Consider then, the disjunction of *e'* with *e''*. This is a logical consequence of *e'*, and since each disjunct in conjunction with *e* fails to justify *h* the disjunction as a whole in conjunction with *e* fails

to justify  $h$ . So, in accordance with (ivd),  $S$  will have to be completely justified in believing the disjunction of  $e'$  with  $e''$  to be false. But, for any disjunction,  $S$  is completely justified in believing it to be false if and only if he is justified in believing, of *each* disjunct, that it is false. We have already assumed that  $S$  is not justified in believing  $e''$  to be false. Examples exhibiting this structure are not at all difficult to find, so we may conclude that the analysis of knowing incorporating (ivd) is too weak—it fails to exclude a case of nonknowledge. In each of these examples, clauses, (1), (2), and (3) of criterion (C) are satisfied. So, the problem in each case has to do with the proposed explication of indefeasible justification, and we may conclude that (ivd) is inadequate.

## II

We can give an intuitive characterization of what is wrong when a man's justification is defeasible by saying that he is less than "ideally situated" with respect to the evidence bearing upon  $h$ . The notion of being *ideally* situated is, of course, a pipe-dream. In any full sense of the term, anyone who falls short of omniscience is thereby less than ideally situated. Nevertheless, we can make perfectly good sense out of the idea that at some times we are better off than at others with respect to the evidence we have for various propositions. In contexts of inquiry and investigation it is normal and natural to speak of the investigator as improving his evidential base. Let us imagine that a man has some evidence  $e$  which is strong enough to justify  $h$ , and let us suppose that over a period of time his evidential position with respect to  $h$  constantly improves (or, approaches some ideally situated state). He comes to know more and more evidence bearing upon  $h$ . It seems intuitively appealing to say that if his justification is indefeasible, then this improvement of his evidential position would never result in a weakening of his (present) justification for  $h$ . The problem is, can we make this intuitively appealing, but vague notion precise?

Recently, Risto Hilpinen<sup>11</sup> and Peter Klein<sup>12</sup> have independently proposed virtually the same analysis of defeasibility, and this analysis is very

much in line with the intuitively appealing idea expressed above. I shall concentrate here on Hilpinen's proposal, but I believe that very similar remarks can be made about Klein's proposal.

In constructing his characterization of defeasibility, Hilpinen refers to a principle put forth by Jaakko Hintikka. Hilpinen calls this the "extendability thesis." This principle states that if a man genuinely knows some proposition  $h$ , then no matter what else he might come to know, he will not lose the knowledge that  $h$  as a result. Hintikka puts the matter in the following way:

If somebody says "I know that  $p$ " in this strong sense of knowledge, he implicitly denies that any further information would have led him to alter his view. He commits himself to the view that he would still persist in saying that he knows  $p$  is true—or at the very least persist in saying that  $p$  is in fact true—even if he knew more than he now knows.<sup>13</sup>

Leaving aside the dubious parts about what men would persist in saying, we can express the extendability thesis as follows:

(ET) If  $S$  knows that  $h$ , then for any true proposition  $q$ ,  $S$  would know  $h$  even if he knew more than he now knows by knowing that  $q$ .

According to (ET), then, the corpus of propositions of which a man has genuine knowledge is always extendable to new instances of knowledge without detriment to any "previous" members of the corpus. Hilpinen suggests that the extendability thesis be adopted as "... a condition of adequacy for the definition of defeasibility."<sup>14</sup> That is, any characterization of defeasibility that we come up with must be such that if knowledge is indefeasibly justified true belief, then any instance of knowledge in accordance with our defeasibility analysis must satisfy the extendability thesis.

Let us consider, however, whether (ET) is a principle that we wish to accept. To begin, (ET) is vaguely formulated, as is often the case when a subjunctive clause is used. We are not told by (ET) what *else* is permitted to happen to  $S$  in the event that he should know more than he now knows by coming to know  $q$ . What does the expression "more

<sup>11</sup> Hilpinen, "Knowledge and Justification," *op. cit.*

<sup>12</sup> Klein, "A Proposed Definition of Propositional Knowledge," *op. cit.* Although Klein does not explicitly characterize his "fourth" condition as a defeasibility condition, I believe that, given the context of his discussion, it is reasonable to interpret his definition as a defeasibility definition.

<sup>13</sup> Jaakko Hintikka, *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca, 1962), pp. 20–21.

<sup>14</sup> Hilpinen, *op. cit.* It should be pointed out that (ET) is not the version of the extendability thesis that Hilpinen uses. For his version, see (ET<sub>3</sub>) and footnote 15 below.

than" in (ET) allow for? We might interpret this as saying that  $S$  knows more at  $t_2$  than he knew at  $t_1$  if and only if at  $t_2$   $S$  knows everything that he knew at  $t_1$  and knows at least one thing that he did not know at  $t_1$ . We can interpret the extendability thesis in accordance with this sense of "knowing more than" as follows:

(ET1) If  $S$  knows that  $h$ , then for any true proposition  $q$ ,  $S$  would know  $h$  even if he knew everything that he now knows and also knew  $q$ .

(ET1) is a trivial analytic truth, however, for it says that if  $S$  knows  $h$ , then he would still know  $h$  even if he continued to know  $h$  and knew something else as well. This is not a very interesting principle. Any analysis of defeasibility would satisfy its demands.

A more interesting way of interpreting "more than" is to say that  $S$  knows more at  $t_2$  than he knew at  $t_1$  if and only if at  $t_2$  the sum total of  $S$ 's information about the world is greater than it was at  $t_1$ . The phrase "sum total of  $S$ 's information" is not intended to express any very technical sense of information. If we assume that some rough measure of the cognitive content of sentences in natural languages is available, then the sum total of  $S$ 's information would simply be the measure assigned to the conjunction of everything that he knows. Clearly, the sum total of  $S$ 's information could increase over a period of time even if the items of information that he has were to change; that is, it is compatible with (S2) that  $S$  should come to know more than he now knows even though he loses some of his present knowledge. Hence, we get the following nontrivial interpretation of the extendability thesis:

(ET2) If  $S$  knows that  $h$ , then for any true proposition  $q$ ,  $S$  would know  $h$  even if the sum total of his information about the world were to increase as a result of his coming to know  $q$ .

It is not difficult to see, however, that (ET2) is too strong. The reason why it is too strong is that it fails to specify any epistemic connection between the supposed increase in (true) information, and the question whether  $S$  would still know  $h$  were that

increase to occur. If I forget something, and also learn a number of new things, then (ET2) gives the result that I never did know the thing I have forgotten, which is absurd. It seems clear that the intent of the extendability thesis is to specify that if a man knows something, then his coming to have more information about the world would not result on epistemic grounds in his losing the previous knowledge. If he does lose the previous knowledge (through mere forgetting, for example) this will be for other reasons. Perhaps we can capture this idea in the following way:

(ET3) If  $S$  knows that  $h$ , then for any true proposition  $q$ , if the sum total of  $S$ 's information about the world were to increase as a result of his coming to know  $q$ , then he would not lose his knowledge that  $h$  as an epistemic result of his coming to know  $q$ .

I suspect that the condition formulated here is pretty close to what Hilpinen has in mind.<sup>15</sup> Unfortunately, even this formulation of the extendability thesis places restrictions on an analysis of knowing that are too strong. We earlier imagined a man who had just thrown a rock at a window, and who was about to suffer sensory paralysis. A careful study of that example will reveal that (ET3) is not satisfied, even though the man knows that the window will break.

Assuming that (ET3), or something like it, is the intended interpretation of the extendability thesis, these criticisms show that we are under no constraints to limit our analysis of defeasibility to one that satisfies the extendability thesis. It is not surprising that the defeasibility condition proposed by Hilpinen (which is virtually the same as the one proposed independently by Klein) turns out to be too strong. The condition proposed by Hilpinen, which demonstrably satisfies the extendability thesis, is this:

(ive)  $S$ 's justification for  $h$  is indefeasible (that is, for any true proposition  $q$ ,  $S$  would be completely justified in believing  $h$ , even if he were completely justified in believing  $q$ ).

<sup>15</sup> Hilpinen derives his defeasibility condition from his formulation of the extendability thesis: "If  $a$  knows that  $h$  on the basis of  $e$ , then for any true proposition  $q$ ,  $a$  would be completely justified in believing that  $h$  even if he knew that  $q$ " (*op. cit.*, p. 31). This is vague in just the way (ET) is. On the basis of personal correspondence, I believe that (ET3) is close to the interpretation of the extendability thesis that Hilpinen wants, although I am not sure he would approve of my addition of the phrase "if the sum total of  $S$ 's information about the world were to increase . . .". Even if that phrase were dropped, the resulting thesis would still be too strong—as would (ive).

The example of the rock thrower shows this to be too strong, just as it shows (ET<sub>3</sub>) to be too strong.

I have construed the intuitive idea behind the approach of Hilpinen and Klein in terms of approaching an ideally situated state with respect to the evidence bearing upon some proposition. From a given epistemic position a man might approach such an ideal state to any of a variety of degrees. He might pick up a bit of information here, and a bit there, in piecemeal fashion. Ironically, a man will sometimes wind up in a position that is worse than the one from which he began, even though he has moved closer to an ideally situated position by acquiring some new information.

The proposals of Hilpinen and Klein fail precisely because they allow for the possibility that a man's justification might be defeated by acquisition of an arbitrary limited portion of the evidence that he does not already possess. Given criterion (C) this leniency proves undesirable. The acquisition of such a limited portion of available new evidence is, we might say, one extreme on the continuum of ways in which a man could move toward a more ideally situated position. The other extreme is the acquisition of all the additional available information, but this extreme is only an epistemologist's pipe dream. We need something in between. In what follows I shall develop the idea that, of the possible ways in which a man might move toward a more ideally situated position with respect to justification, only some of those ways are relevant to the question whether his justification is defeasible.

### III

When we talk about what would happen if a man were to become more ideally situated with respect to the evidence bearing upon some hypothesis, we are talking about how things might have been, or might come to be otherwise epistemically for him. I find it fruitful to characterize this kind of situation in terms of what I have elsewhere called *epistemic frameworks* and *alternatives* to epistemic frameworks.<sup>16</sup> An epistemic framework (abbreviated "Fs") is a set of epistemic descriptions of the forms "*S* believes that *p*," "*S* knows that *p*," "*S* is justified in believing

that *p*," and so forth, such that the set completely describes the epistemic state of affairs of the person *S* at a given time. An alternative to an epistemic framework is another set of epistemic descriptions whose membership is determined in one of a variety of ways relative to the membership of the epistemic framework to which it is an alternative. In the interesting cases, a defined type of alternative to an epistemic framework will characterize an epistemic state of affairs of the person *S* which is *possible* but in fact quite different from the state of affairs actually described by his epistemic framework.

There are many subsets of an epistemic framework that are of interest. One of these is the subset consisting of all true expressions of the form "*S* is justified in believing —," and other expressions that characterize *S*'s evidential situation. This set of expressions corresponds to what is usually called *S*'s evidential base, or body of evidence; accordingly, I shall call this subset the *evidence component* of *S*'s epistemic framework. The question whether an epistemic description of the form "*S* is justified in believing —" is a member of *S*'s evidence component is to be decided by proper explication of the concept of justification. Such an explication will yield a set of rules governing the admissibility of these expressions into an epistemic framework. These rules are called rules of epistemic inference. A number of plausible rules of epistemic inference have been suggested,<sup>17</sup> but a great deal of work needs to be done in this difficult area.

Assuming the availability of an adequate set of epistemic inference rules, we can specify the conditions under which a man, *S*, is justified in believing something as follows:

*S* is justified in believing that *h* if and only if the expression "*S* is justified in believing that *h*" is derivable from the other members of *S*'s evidence component in accordance with the rules of epistemic inference.

Let us suppose, then, that our subject *S* is justified in believing that *h*. We are concerned with the question whether his justification is indefeasible. This question can now be recast in the following way. Suppose we have defined a type of alternative to *S*'s epistemic framework such that if that alterna-

<sup>16</sup> Marshall Swain, "The Consistency of Rational Belief," in *Induction, Acceptance, and Rational Belief*, *op. cit.*, pp. 27-32, and "An Alternative Analysis of Knowing," *op. cit.*

<sup>17</sup> For some remarks about epistemic inference rules see the two papers referred to in footnote 15. See also Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, chs. 2 and 3, and "On the Nature of Empirical Evidence" in Chisholm, *et. al.*, *Theory and Experience*. See also Herbert Heidelberger, "Chisholm's Epistemic Principles," *Nous*, vol. 3 (1969), pp. 73-82.

335085

tive had obtained (or will obtain) then  $S$  would be more ideally situated evidentially than he in fact is; relative to the evidence component of such an alternative would  $S$  still be justified in believing that  $h$ ? If so, then his justification is indefeasible; if not, then his justification is defeasible. The critical question is, can we define a type of alternative to an epistemic framework such that the defeasibility or indefeasibility of a man's justification rests upon what would be the case if that alternative were actualized?

To see what kind of alternative we need, we can reflect again on the examples of defeasible and indefeasible justification raised earlier. In some of those examples, such as the rock-thrower example, even though there is "available" some negative counterevidence, the defeating effect of this counter-evidence is itself overridden. In the other cases the defeating effect of the counterevidence is not overridden, and the man's justification is defeated. In all of these examples, where there is some negative counterevidence it consists of some proposition or set of propositions that the person  $S$  is *in fact* justified in believing to be *false*. The wedding ceremony example shows us, indeed, that we must hold it to be a necessary condition of a proposition's being a defeating proposition that it be one the person  $S$  is justified in believing to be false. If we do not impose this requirement we allow for the possibility that a man's justification can be overridden by some bit of evidence that is available as a result of accident. In most cognitive situations, we are justified in believing a number of false propositions along with the true propositions that we are justified in believing. Hence, one way in which a man might approach an ideally situated evidential position is that it should become evident for him that some (that is, at least one) of the false propositions that he is justified in believing to be true are in fact false. We could define a type of alternative to a man's epistemic framework which describes a possible situation of this sort. To make defeasibility dependent upon what would happen if this kind of alternative were actualized would not, however, differ essentially from the proposals of Hilpinen and Lehrer and Paxson. Our earlier discussion of those proposals shows that this type of alternative will not serve our purposes.

A more promising approach is to consider what would happen if a man were to become justified in believing to be true *all* of those true propositions that he is in fact justified in believing to be false, and where the remainder of his epistemic situation

changes only in some minimal way required to preserve consistency. We can specify what such a situation would look like by defining a type of alternative, which I shall call an *evidence-restricted* alternative:

- (D8)  $Fs^*$  is an evidence-restricted alternative to an epistemic framework  $Fs$  if and only if
- (i) for every true proposition  $q$  such that " $S$  is justified in believing not- $q$ " is a member of the evidence component of  $Fs$ , " $S$  is justified in believing  $q$ " is a member of the evidence component of  $Fs^*$ ,
  - (ii) for some subset  $C$  of members of  $Fs$  such that  $C$  is maximally consistent epistemically with the members generated in (i), every member of  $C$  is a member of  $Fs^*$ , and
  - (iii) no other propositions are members of  $Fs^*$  except those that are implied epistemically by the members generated in (i) and (ii).

In this definition, there are two technical expressions that need to be explained. The expression "is epistemically consistent with" is to be understood as follows: Relative to a given set of rules of epistemic inference, two epistemic descriptions  $E$  and  $E'$  are epistemically consistent just in case neither of them implies epistemically the denial of the other. The notion of maximal consistency can be defined in this context in the following way: A subset  $C$  of members of  $Fs$  is maximally consistent epistemically with the members generated by (i) if and only if (a) the union of the set of members generated by (i) with the members of  $C$  is epistemically consistent and (b) for any other subset  $C'$  of  $Fs$  such that  $C$  is a proper subset of  $C'$ , the union of the set of members generated by (i) with  $C'$  is epistemically inconsistent.

An evidence-restricted alternative is, then, a complete description of how things might change for  $S$  if his body of evidence were purged of all false members, each of these being replaced by its denial. Such an alternative describes one very carefully circumscribed way in which a man's epistemic position can improve. We can make defeasibility dependent upon what would happen if one's epistemic position were to improve in just this carefully defined way by replacing our earlier condition (ive) with:

- (ivf)  $S$ 's justification for  $h$  is indefeasible (that is, there is an evidence-restricted alternative  $Fs^*$  to  $S$ 's epistemic framework  $Fs$

such that "S is justified in believing that *h*" is epistemically derivable from the other members of the evidence component of *Fs*\*).

If (ivf) is used to construct an analysis of knowing, then the result will avoid the various problems we have considered so far. I shall consider two cases in illustration of the merits of (ivf).

First, let us consider one of the cases proposed by Gettier. Suppose that Jones has recently been told by his employer that he will be the next vice-president. The employer has in the past been honest and reliable. Moreover, suppose Jones knows that he has ten coins in his pocket, having just counted them. These facts plus background evidence constitute his evidence *e*. Jones recognizes that his evidence *e* justifies him in believing that *q*, namely, "Jones has ten coins in his pocket and will be the next vice-president." He also recognizes that *q* entails *h*, namely, "The man who will be the next vice-president has ten coins in his pocket." He concludes that *h* is justified for him since *h* is entailed by the justified proposition *q*. Moreover, *h* is true, but not for the reasons that Jones thinks it is true. The employer has untypically made a mistake. It is Smith, not Jones, who will be the next vice-president, and Smith coincidentally has ten coins in his pocket. So, Jones does not know that *h* even though he has justified true belief. And, the reason he does not know *h* is that his justification is defeasible. If Jones were to become ideally situated in the manner specified in the definition of an evidence-restricted alternative, it would no longer be the case that *h* is justified for him, relative to the other members of his evidence component. Let us see why this is so. In the example, there are a number of false propositions that Jones is justified in believing. Some crucial ones are:

- (1) The employer has spoken the truth.
- (2) Jones will be the next vice-president.

According to the definition of an evidence-restricted alternative, we first list as evident for Jones the denials of all those false propositions that Jones is in fact justified in believing to be true. Thus, among the members of the evidence component of the alternative will be:

- (1') Jones is justified in believing that the employer has not spoken truly.
- (2') Jones is justified in believing that he will not be the next vice-president.

The next step in the formula is to add to the alternative all those statements in some subset *C* of members of *Fs* that is maximally consistent with the members generated by the first step. One subset will include the following:

- (3) Jones is justified in believing that the employer told him he would be the next vice-president.
- (4) Jones is justified in believing that he has ten coins in his pocket.
- (5) Jones is justified in believing that the employer has in the past been honest and reliable.

However, since (1') and (2') are already members of the alternative, we cannot consistently add that (1) and (2) are justified for Jones. It is quite clear that given the membership of the alternative evidence component in this example, the proposition that the man who will get the job has ten coins in his pocket is not justified for Jones under that alternative. Moreover, there does not appear to be any other maximally consistent set of members of *Fs*, relative to (1') and (2'), such that *S* will be justified in believing *h* relative to the evidence-restricted alternative including that set. Hence, Jones's justification is defeasible by (ivf) and he does not know *h*. This is the desired result.

Let us now consider the rock-thrower example used against Hilpinen's proposal above. It will be recalled that this is a case where we would say *S* knows that *h*, but where the other conceptions of defeasibility give us the wrong result. I shall now argue that *S*'s justification is indefeasible in this example, according to (ivf).

In the rock-thrower case, as in the previous example, there are a number of false propositions that *S* is justified in believing. The important members of *S*'s epistemic framework in this regard are:

- (6) *S* is justified in believing that he will see the window break.
- (7) *S* is justified in believing that he will hear the window break.
- (8) *S* is justified in believing that his sensory mechanisms will continue to function properly.

Since each of these justified things is false, we first include the following in the evidence-restricted alternative:

- (6') *S* is justified in believing that he will not see the window break.

- (7') *S* is justified in believing that he will not hear the window break.  
 (8') *S* is justified in believing that his sensory mechanisms will not continue to function properly.

We then add to the alternative some subset of members of *S*'s epistemic framework that are maximally consistent with (6')–(8'). One subset will include virtually all of *S*'s *original* evidence, such as:

- (9) *S* is justified in believing that the rock will hit the window.  
 (10) *S* is justified in believing that the window is a normal window.

but will not include (6), (7), or (8). It is clear that in the alternative so generated, it will still be the case that *S* is justified in believing that the window will break. For, even though the evidence component of the alternative contains the potentially defeating (6') and (7'), it also contains (8'), which nullifies the defeating effect of those other members. Hence, we can conclude that *S*'s justification here is indefeasible, and that he has knowledge. Again, we get the desired result.

The proposal embodied in (ivf) is essentially the same as one that I originally suggested in my paper "An Alternative Analysis of Knowing."<sup>18</sup> The examples just considered and the ones considered in the earlier paper show that an analysis of knowing incorporating this proposal is capable of dealing with many of the problems that have been raised in the literature. However, I now believe that (ivf) is inaccurate, for it fails to take into account an important aspect of the notion of defeasibility. I suggested above that the difference between the Gettier case and the case of the rock thrower is that in that latter *S*'s justification is indefeasible because the defeating effect of the available counterevidence is in turn overridden, whereas in the Gettier cases this is not so. These facts become clear when we consider the relevant evidence-restricted alternatives. Unfortunately, I have now found variations on the Gettier cases wherein the defeating effect of the available counterevidence is also overridden, according to (ivf), even though *S* does not have knowledge. All that (ivf) requires for indefeasibility is that in some evidence-restricted alternative it be true that *S* is justified in believing *h*—there is no specification

that his justification in the alternative be the *same* as the justification that he actually has. But it is clearly essential that this be so. Consider the following variation on the Gettier case above. Suppose everything is as before, except that for some reason Jones happens to know that Smith (the man who is really going to get the job) has ten coins in his pocket. If so, then in the relevant evidence-restricted alternative to Jones's epistemic framework it will be evident for Jones that Smith will get the job (since it is *in fact* evident for him that Smith will not get the job) and *also* evident for him that Smith has ten coins in his pocket. But this renders it evident for him (in the alternative) that the man who will get the job has ten coins in his pocket. According to (ivf), Jones's justification for *h* would be indefeasible, and he would have knowledge. But neither of these things is so.

This argument against (ivf) brings out the fact that it is always a particular justification that is defeasible or indefeasible. The fact that if a man were to become more ideally situated in the manner specified by the notion of an evidence-restricted alternative then he might have some *other* justification for *h* is irrelevant to the question whether *the* justification he in fact has is defeasible or not. We need to require that a justification be indefeasible only if that justification *itself* would continue to hold under the process of becoming more ideally situated epistemically. In the rock-thrower case it is clear that even if the man became more ideally situated by coming to know that he would suffer sensory paralysis, etc., the justification that he has for believing that the window will break would remain essentially the same. But, in the revised Gettier case just considered the justification *S* has would be replaced by another in the evidence-restricted alternative.

We must be cautious, however, about requiring that a justification be *exactly* preserved as we move to some evidence-restricted alternative if that justification is to be indefeasible. We shall have to allow for some slight changes here and there. First, it has been effectively argued in the literature<sup>19</sup> that sometimes a man's justification can involve some false evidence, provided this false evidence is not central to his justification. A man might, for example, have two evidential beliefs, each of which is sufficient to justify belief that *h*, but one of which is false. Provided that he fully bases his belief that

<sup>18</sup> See "An Alternative Analysis of Knowing," *op. cit.*

<sup>19</sup> See Keith Lehrer, "Knowledge, Truth, and Evidence," *op. cit.* See also Roderick Chisholm, "The Nature of Empirical Evidence," *op. cit.*, pp. 119–122.



$h$  on both bits of evidence, there is no *prima facie* reason why his justification could not be indefeasible. In such a case, however, when we move to the evidence-restricted alternative the false evidence that he has will no longer form part of his justification. Secondly, in many cognitive situations, particularly those in which a man's justification is indefeasible, the justification that he would have under the alternative will be augmented by new information. Hence, we can say that his justification changes when we move to the alternative; but notice that it does not change in the sense of being *replaced* by a new line of justification.

To rule out the latest counterexample while allowing for the facts just discussed, I shall revise (ivf) as follows:

- (ivg)  $S$ 's justification for  $h$  is indefeasible (that is, there is an evidence-restricted alternative  $Fs^*$  to  $S$ 's epistemic framework  $Fs$  such that (i) " $S$  is justified in believing that  $h$ " is epistemically derivable from the other members of the evidence component of  $Fs^*$  and (ii) there is some subset of members of the evidence component of  $Fs^*$  such that (a) the members of this subset are also members of the evidence component of  $Fs$  and (b) " $S$  is justified in believing that  $h$ " is epistemically derivable from the members of this subset).

In the revised Gettier case above, even though  $S$  is justified in believing that  $h$  under the evidence-restricted alternative, there is no part of the justification that he in fact has which "carries over" to the alternative and which is alone sufficient to

justify  $h$ . This, I believe, is the force of saying that his justification is replaced by a new one in the alternative. Hence, by clause (ii) of (ivg) this example is eliminated.

However, (ivg) allows for the desired kind of changes in  $S$ 's justification when we move to the evidence-restricted alternative. It is compatible with clause (ii) of (ivg) that any false parts of the justification that  $S$  in fact has will be eliminated in the alternative—part (b) of clause (ii) requires that that part of his justification that *does* carry over to the alternative must be sufficient to justify  $h$ . This is the force of saying that the central part of his justification is preserved in moving to the alternative. It is also compatible with clause (ii) that  $S$ 's justification changes under the alternative in the sense of being augmented by new information. Even though clause (ii) requires that the subset which "carries over" be sufficient to justify  $h$ , there is no stipulation that this subset constitutes the total body of justifying evidence that  $S$  has under the alternative. Of course, any additional evidence that  $S$  has in the alternative must be such that clause (i) of (ivg) is satisfied.

I submit that (ivg) is, therefore, adequate as an analysis of epistemic defeasibility. My argument for this claim rests heavily on the criterion (C) introduced at the beginning of this paper. It is perhaps a bit hasty to rest such a large conclusion on the fact that the proposal works in one limited area, namely, the problem of analyzing knowledge. The real test of my proposal will be whether it also works in the other areas where epistemic defeasibility plays a role.<sup>20</sup>

Ohio State University

Received January 2, 1973

<sup>20</sup> A short version of this paper was read at the Eastern Division meetings of the American Philosophical Association, December 29, 1972. Ernest Sosa was the commentator. I am indebted to Sosa, and to a referee of this journal for helpful suggestions. In his commentary, Sosa argued that (DB) needs revision because, given (ivg), it has the result that if  $S$ 's evidence  $E$  renders some false statement  $F$  evident, then in the evidence-restricted alternative it will be evident that not- $F$ , which is incompatible with it's also being evident in the alternative that  $E$ , which in turn implies that  $E$  is not in fact known by  $S$ . This problem can be dealt with by revising (DB) to guarantee that in the alternative there is some  $R$  such that  $R$  is evident for  $S$  and such that  $R$  confers evidence upon not- $F$  as well as upon  $E$ . I regret that space limitations prohibit my giving a more detailed response to Sosa's problem at this time.

### III. THE ABOUTNESS OF EMOTIONS

ROBERT M. GORDON

I SHALL attempt to show that when someone is, e.g., angry about something, the events or states that conjointly are causing him to be angry conform to a certain structure, and that from the causal structure underlying his anger it is possible to "read out" what he is angry about. In this respect, and even in some of the details of the structure, my analysis of being angry about something resembles the belief-want analysis of intentional action.<sup>1</sup>

The intended effect of the proposed analysis is, in part, to enrich our understanding of the cognitive and attitudinal factors in emotion: for the chief elements of the causal structure I shall describe are a belief and an attitude so related in content as to constitute either a wish-frustration (in the case of negative emotions) or a wish-satisfaction (in the case of positive emotions). In part, however, the effect of the analysis is reductionistic. For it makes otiose, in those cases for which it is a correct analysis, the mysterious non-causal relation between an emotion and its "object" which is invoked by the majority of philosophers now writing on emotions.<sup>2</sup>

#### I. CIRCUMSCRIPTION OF THE TOPIC

My intention is to analyze being  $\phi$  about something, where ' $\phi$ ' is replaced by any of the following expressions, among others: "angry," "annoyed," "ashamed," "delighted," "disappointed," "disgusted," "embarrassed," "furious," "glad," "grateful," "horrified," "indignant," "pleased," "resentful," "sad," "sorry," "unhappy," "upset."

If the reader finds the analysis satisfactory for the expressions I have listed, then he is welcome to apply it to any other expression that seems to fit the

analysis as well as these do. I extend a similar invitation with regard to *prepositions*. If, as it appears, the analysis works as well, e.g., for being angry *at* something as for being angry *about* something, for being upset *over* something as for being upset *about* something, and so on, all the better. I can even think of some adjectives that seem to fit the analysis, though they do not, or do not naturally, accept the preposition "about": e.g., "grateful (for)," "proud (of)." What does appear to hold true of all viable replacements for ' $\phi$ ' is that they may be followed by "that"-clauses; and when so followed, the truth of the resulting sentence presupposes the truth of the "that"-clause, i.e., of the sentence that has been transformed into a "that"-clause. This, as will emerge in Sect. V, is because the concept of *knowledge* plays an essential role in the analysis.

I shall not pretend that the analysis will work for every state we are wont to call an emotion, nor indeed that it will work only for such a state; nor do I intend the analysis to cover everything that might be counted an "object," e.g., of anger: the person one is angry with, the person one vents one's anger on, and so forth. By this disclaimer, I shall avoid a trap that several writers seem to have fallen into. Anthony Kenny, the most influential of recent philosophers of emotion, reasons essentially as follows (I paraphrase for brevity): "There are many cases in which it is natural to think of the object of emotion as its cause: e.g., 'I was angry because he burst in without knocking,' and, 'Her behavior made me most embarrassed.' But there are other examples that show that the relation of an emotion to its object cannot be one of effect to cause: e.g., 'I dread the next war,' 'I hope Eclipse will win.' So, despite appearances, the ostensibly

<sup>1</sup> As presented in Donald Davidson, "Actions, Reasons, and Causes," *The Journal of Philosophy*, vol. 60 (1963), pp. 685-700, and Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, 1970).

<sup>2</sup> Among them, Anthony Kenny, in *Action, Emotion, and Will* (London, 1963); D. F. Pears in "Causes and Objects of Some Feelings and Psychological Reactions," *Ratio*, vol. 4 (1962), pp. 91-111; George Pitcher, in "Emotion," *Mind*, vol. 74 (1965), pp. 326-346; Irving Thalberg, in "Emotion and Thought," *American Philosophical Quarterly*, vol. 1 (1964), pp. 45-55; and B. A. O. Williams in "Pleasure and Belief," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXIII (1959), pp. 57-72.

causal expressions in the earlier examples must be understood as having a special non-causal sense."<sup>3</sup>

But if the two sets of examples show anything, it is that it is a false start even to speak of "*the* relation between an emotion and its object." What is missing from Kenny's reasoning is some reason for thinking that, despite appearances, the same "relation" holds, on the one hand, between one's anger and the remark that makes one angry, and, on the other hand, between one's dread and the yet-to-come war that one dreads. Since I do not claim to be analyzing "*the* relation between an emotion and its object," my analysis is not threatened by alleged counterexamples advanced on the grounds that "here is an object of emotion that does not fit the analysis."

There is another limitation on the scope of the proposed analysis. There is a rich variety of grammatical constructions that may be tacked onto the  $\phi$ -words listed at the head of this section. These seem to designate an even greater variety of entities: facts, possibilities, properties, events, concrete objects, and so on. This analysis will concern, specifically, being angry (etc.) about a *fact*. I do not think that anything is lost by this limitation, however. For I am convinced that it is facts that we are primarily angry about. Indeed, I should like to say (but for the air of paradox it would convey) that, strictly speaking, it is only facts that we are angry (etc.) about. Any linguistic forms that may suggest otherwise can be explained, I should claim, on this hypothesis.

In this I am helped by some recent work in linguistics. According to a hypothesis of Kiparsky and Kiparsky,<sup>4</sup> the following sentences should have a common deep-structural explanation: "Jules is angry about the fact that Jim has phoned Jane," "Jules is angry about Jim's having phoned Jane," and "Jules is angry that Jim has phoned Jane." Of these three sentences the one that best reflects their common deep structure is the first; the others are derivable by various optional transformations, one of which deletes the explicit mention of a fact. To this we may add that after a  $\phi$ -word almost any form of noun phrase demands a sentential, and hence in this context a fact-like interpretation. For example: "the rudeness with which he was treated" = "the fact that he was treated (so) rudely"; "the way they treated him" = "the fact that they treated him the way they (in fact) did." Concrete nouns do not convey enough information

to yield a unique fact-like interpretation. Nonetheless, I should claim that there is always a fact (or putative fact) in the offing, which the context of utterance usually enables one to pick out. Thus, if Smith is angry about *his lawn*, there is, as we might put it, *something about* his lawn that he is angry about: e.g., the fact that his lawn is bare, or the fact that his lawn looks the way it (in fact) does.

If it is true that one is primarily angry about facts, then it becomes possible to analyze being angry about something in terms of the "propositional attitudes" that give rise to one's anger. This I shall now proceed to do.

## II. THE BELIEF REQUIREMENT

My point of departure is the following two assumptions:

- (1) If *S* is angry (etc.) about the fact that *p*, then *S* is angry.
- (2) If *S* is angry (etc.) about the fact that *p*, then *S* believes that *p*.

Assumption (1) requires some explanation. With any of the  $\phi$ -words listed, "is  $\phi$ ," "appears  $\phi$ ," and "feels  $\phi$ " are grammatically complete verb phrases taking only optional complementation by a prepositional phrase or a "that"-clause. Thus, "Jules is (or: appears, feels) angry" is a sentence. This is not to deny that, at least in standard cases, if someone is angry then he is angry about something. Perhaps "is angry," like "is a father," is a relational predicate that deletes one of its arguments. All of this I understand to be consistent with assumption (1) above.

In fact, as I explain in Sect. VI, I do not think it just a contingent fact that emotions are generally *about* something, and that particular emotion-types such as anger are about particular kinds of things. This may be one way in which emotions are distinct from sensations, contrary to what Hume and James seem to have held. But there are other differences. Emotions, unlike sensations, seem to me best treated as either dispositional or functional states defined, at least in part, by the phenomena they are capable of explaining. I regard being angry, for example, as a state that becomes manifest in a variety of ways, the most important being the acquisition of desires of a hostile or aggressive sort. With other emotional states, such as being upset, the chief manifestation may be interference

<sup>3</sup>Kenny, *op. cit.*, pp. 71-75 *et passim*.

<sup>4</sup>Paul Kiparsky and Carol Kiparsky, "Fact," in Bierwisch and Heidolph, *Progress in Linguistics* (The Hague, 1970).

with thought and action; with others still, it may consist in physiological disturbances and the correlate bodily sensations.

Given assumptions (1) and (2) above, if  $S$  is angry about the fact that  $p$ , then  $S$  believes that  $p$  and is angry. Must it be true, in addition, that  $S$ 's believing that  $p$  is related in a certain way to  $S$ 's being angry? Or is it possible that these two conditions, together with some third condition yet to be mentioned, simply "add up" to  $S$ 's being angry about the fact that  $p$ —with no requirement that the conditions be interrelated in a certain way? The second alternative is unacceptable. This becomes obvious when we consider an example in which a person is independently angry about two or more facts at once; this, together with the supporting discussion in Sect. III, seems to me to place the matter beyond doubt.

Suppose that Junior has tampered with Daddy's camera; he has opened the back, fogging the film and ruining 20 irreplaceable shots of Hongkong. Daddy, as he comes home to this scene, happens already to be fuming over the fact that his car will be laid up at the repair shop for another day. He learns about Junior's misdeed. So now he believes that Junior has ruined the Hongkong shots, and, at the same time, he happens to be angry. But is he angry about the fact that Junior has ruined the Hongkong shots? His happening to have been angry just then seems irrelevant to this question. If it remains a mere coincidence that he believes that Junior ruined the pictures and at the same time is angry (because he is angry about his car's being laid up), he is no closer to being angry about Junior's having ruined the pictures than if he had not been angry at all. No matter what further conditions we suppose him to satisfy, he will not be angry about what Junior did. His having the belief and his being angry must be related in such a way that it is not a mere coincidence that they are co-present now.

What is required is that Daddy's anger acquire a new footing in his belief that Junior has ruined the Hongkong shots; or, to complement that metaphor with others, that it gain new sustenance from that source, and hence a new lease on life. The point of these metaphors is that, whereas formerly he would not have been angry were it not for the fact (or his belief) that his car would be laid up, this is no longer true; for the new belief—viz., that Junior has ruined the Hongkong shots—is sufficient, under the circumstances, to make or to keep him angry. Thus it is not a mere coincidence that

on the one hand he believes that Junior has ruined the shots and on the other hand he is angry.

I have overstated the case somewhat. For one thing, it appears to be a datum of psychology that anger sometimes begets anger: more precisely, in a sense to be explained in Sect. III, it is sometimes a catalyst for further anger. Thus, had Daddy not been raised to a sufficient pitch of anger by his belief (a) that his car would be laid up, it is possible that his belief (b) that Junior had ruined the Hongkong shots would not have been sufficient, under the remaining circumstances, to make or to keep him angry. In such a case, he would not have been angry at all, had he not also believed (a). So what is crucial in the above example is not *that belief (a) is no longer necessary*, i.e., that Daddy might have been angry, under the circumstances, even without that belief; but rather, simply, *that belief (b) is sufficient*, under the circumstances, to make or to keep him angry. We can show this more directly by supposing that, through some bizarre connection, Daddy would not have cared about Junior's ruining the Hongkong shots, but for the fact that his car would be laid up. (A more perspicuous example of this sort is given in Sect. IV: Sam is angry about having to go to work at 6 A.M. and also angry about his neighbor's late-evening violin-playing. But he wouldn't have cared about the violin-playing, except that he had to go to work at 6 A.M. The belief that he had to go to work at six thus remains necessary in either case, violin-playing or no violin-playing.)

My position can now be generalized. There is a Belief Requirement (BR) for being angry about a fact, as follows:

(BR) If  $S$  is angry about the fact that  $p$ , then  $S$ 's believing that  $p$  is sufficient for  $S$  to be angry, given some existing conditions that are not themselves sufficient for  $S$  to be angry.

Note that BR stipulates only that there exist *some* conditions  $c$  such that  $c$  conjointly with  $S$ 's belief suffice for  $S$  to be angry, where  $c$  alone do not suffice. There might exist *other* conditions, however, that do suffice for  $S$  to be angry, even apart from  $S$ 's belief that  $p$ . As to the notion of "sufficiency," I understand it to have the force of the "if-then" of subjunctive conditionals. Thus, not only is it true that, whenever  $c$  obtains and  $S$  believes that  $p$ , then  $S$  is angry; further, whenever  $c$  obtains and  $S$  does not believe that  $p$ , the following subjunctive con-

ditional will be true: "If *S* were to believe that *p*, then *S* would be angry."

(Perhaps it is worth mentioning that I do not regard (BR) as committing me to determinism in any form. Should someone maintain that there is an element of *assent* that is involved whenever a person becomes angry, and that assent is an act, and that acts do not have causally sufficient conditions, I need not dispute with him here. Obviously such assent is not of itself *sufficient* for anger. If it is *necessary*, then we must number it among the conditions *c*: i.e., among the conditions which, though alone they do not suffice for *S*'s anger, do suffice in conjunction with *S*'s belief that *p*. Even if such assent did not itself have causally sufficient conditions, this would not show that anger does not have causally sufficient conditions.)

### III. SOME A PRIORI LIMITS OF PSYCHOPHARMACOLOGY

The force of BR can be shown by the following example. With a sufficient dosage of the right drug, it would seem possible to put any subject into a state of depression or a state of elation. I shall assume for now that a state of being angry can be induced in the same way. (As we shall see in Sect. VI, BR makes such an assumption doubtful, if being angry requires being angry about something.) Suppose that *S*, a subject in a psychological experiment, has been filling out a tiresome and intrusive questionnaire, and that while doing this *S* has grown angry. But suppose all that is causing *S* to be angry is a drug recently introduced into *S*'s bloodstream by injection; and that, in particular, no belief of *S*'s bears any causal relation to *S*'s being angry. *S* is aware that he is angry but is not aware that his being angry has anything to do with the drug in his bloodstream. (Either he is unaware that he was given an injection, or he was led to believe that the drug injected had a quite different effect.) *S* takes himself to be angry about the questionnaire he is required to fill out—more specifically, about the fact that he is required to fill out so tiresome and intrusive a questionnaire.<sup>5</sup>

The problem is to decide whether, under these circumstances, *S* is actually angry about the questionnaire, or indeed whether *S* is angry about anything at all. It would appear that BR requires that

both questions be answered in the negative. But before applying BR to this example, we need to sharpen the distinction between the drug described and other possible drugs.

The drug described above is to be distinguished from a drug which, though incapable of inducing anger on its own, increases the irritability of the subject so as to make him susceptible to anger at things that otherwise would not have made him angry: e.g., things too trivial or too remote, such as minor slights and distant wars. Such a drug would have to work in complicity with the subject's beliefs, for by itself it does no more than lower the threshold of anger. I shall call such a drug an *anger catalyst*. (As we saw in the preceding section, anger itself may act as an anger catalyst, in this sense.) In contrast, a drug of the sort described above, i.e., one that induces anger without the help of any belief, I shall call an *anger agent*.

For completeness' sake, we should also distinguish between an anger agent and a belief-inducing drug, i.e., a *belief agent*. For a belief agent may on occasion induce a belief that suffices, under the circumstances, to make a person angry. Like an anger agent (and unlike an anger catalyst) it does not work *in conjunction with* the subject's beliefs. But unlike an anger agent it does not induce anger directly: that is, it induces anger only by inducing a belief.

In our original experiment the subject was administered an anger agent. It was further stipulated that the drug (together, of course, with the required background conditions) was *all* that was causing *S* to be angry: this was meant to rule out causal overdetermination. Thus there was no belief of *S*'s that also sufficed, under the circumstances, for *S* to be angry. According to BR, therefore, *S* was not angry about the questionnaire. *S* was not angry *about* anything at all.

Although my intuitions regarding the experiment match this result, it is difficult for me to estimate to what degree they have been shaped by a prior commitment to BR. In fairness, I should say that I have at times felt some inclination to ask, "Isn't what *S* is angry about just a matter of *how S feels*, and isn't this entirely separate from the question of what is causing *S* to feel this way?" To cut what I think is the heart of the problem, what stands behind this objection is the principle that

<sup>5</sup> This "experiment" was inspired by the epinephrine experiments described in Stanley Schachter and Jerome E. Singer, "Cognitive, Social, and Physiological Determinants of Emotional State," *Psychological Review*, vol. 69 (1962), pp. 379-399. In the epinephrine experiments, however, what is induced by the drug is not anger but mere physiological arousal, according to the authors.

what a person sincerely reports himself angry about is the final arbiter of what he is presently angry about; or the related principle that what a person is angry about is entirely a matter of how he feels, in a sense in which to feel a certain way and to seem (to oneself) to feel that way are indistinguishable states. After two preliminary skirmishes, I shall offer what I consider to be two telling objections to the first of these principles. Appropriately modified, they apply also to the second.

First, people misreport that they are angry about "the fact" that something is so, when in fact it is not so. Jules is *not* angry about "the fact that Jim phoned Jean" if it is not a fact that Jim phoned Jean, no matter what Jules may report. (As my remarks in Sect. VI should suggest, there is more to this objection than meets the eye.)

Secondly, many people do not like to admit to themselves that they can get angry about certain things. A father, unwilling to admit that he is angry about his daughter's intention to marry, may prefer to see himself as being angry over her particular choice of a marriage partner. Assuming that he would have been angry no matter whom she had chosen, so long as she intended to marry someone, BR leads to what again seems the right conclusion, namely that he is angry about her intending to marry rather than about her particular choice of partner. Whatever we think of psycho-analytic examples in general, it is hard to deny that one's beliefs as to what one is angry about may, like other beliefs, be responsive to one's wishes.

The following objections, I think, are decisive. First, if someone's sincere report can be shown to be based on ignorance, it loses any claim it may otherwise have had to being the final arbiter of its own truth.<sup>6</sup> Suppose that the subject in our experiment would not have taken himself to be angry about the questionnaire, had he known that it was the drug, and not any of his beliefs, that was causing him to be angry. It is thanks to his ignorance of the true cause of his anger that he reports himself to be angry about the questionnaire. This not only would undermine the report's claim to being the final arbiter but would spoil its value even as *evidence* of its own truth.

Finally, the first-person report is only one of many ways in which one's anger may make itself manifest to others. Anger may also be displayed in facial expression, bodily set, tone of voice, action, and

manner of action, as well as in ways less easily discernible. Any of these, and especially action, may also help to indicate what one is angry about. In addition to the manifestations of anger, one of the chief indicators of what a person is angry about is what he has reason to be angry about (or what he has reason to be angry about, given his attitudes). Where any of these indicators fails to square with the first-person report, then there is reason to question that report, sincere or not sincere.

I think I have said as much as I need to say in support of BR in particular; and I rest my case, not on any single example, but on the entire discussion. Further support will come from showing how BR contributes to the analysis of being angry about a fact. For if the analysis proves on the whole to be a plausible and a useful one, this consideration should overcome any minor qualms the reader may have regarding one or two of the examples offered in support of BR.

It is obvious that BR alone does not provide a complete analysis of being angry about a fact. But together with another requirement to be introduced in the next section, it does constitute the *basic* analysis not only of being angry about a fact but also, *mutatis mutandis*, of being upset, pleased, or sorry (etc.) about a fact.

#### IV. HOW TO FIND OUT WHAT YOU ARE ANGRY ABOUT: THE HARD WAY

The predicates listed at the head of Sect. I can readily be segregated into two lots: namely, those that designate *positive* states and those that designate *negative* states. *Positive*: "delighted," "glad," "grateful," "pleased." *Negative*: "angry," "annoyed," "ashamed," "disappointed," "disgusted," "embarrassed," "furious," "horrified," "indignant," "resentful," "sad," "sorry," "unhappy," "upset."

When *S* is delighted (about the fact) that *p*, *S* has a *pro*-attitude toward its being the case that *p*. The same is true for any of the predicates on the "positive" list. When *S* is angry (about the fact) that *p*, *S* has a *con*-attitude toward its being the case that *p*. The same is true for any of the predicates on the "negative" list. Some care is required in describing these attitudes. We should not speak here of *wanting* or *desiring* something to be the case. One reason is that when *S* is delighted or angry

<sup>6</sup> We should of course distinguish between a report that is based on ignorance and a report of a state that is based on ignorance: e.g., if I report that I believe that Aristotle was born in Athens, and my belief is based on false assumptions, my report is not undermined.

about the fact that  $p$ ,  $S$  already knows it to be the case that  $p$ . This leaves  $S$  with no possibility of instrumental or aversive action. In such a case the notions of wanting and desiring are usually thought inapplicable. Most obviously, where the object phrase is in the past tense the verb "want" gives us a deviant utterance: e.g., "I want Junior not to have ruined the Hongkong shots."

For this and other reasons it is preferable to speak of "wishing" something to be so. Thus, if Daddy is pleased that Junior ruined the shots, then he wishes him to have ruined the shots; if Daddy is angry or upset that Junior ruined the shots, then he wishes him *not* to have ruined the shots. I use the accusative-and-infinitive construction after "wish" to avoid the subjunctive mood, which in the past tense would presuppose that what is wished is contrary to fact. It will be convenient to introduce conventional formulas, as follows: When  $S$  is pleased (etc.) about the fact that  $p$ , then  $S$  *wishes it to be the case that  $p$* ; when  $S$  is angry (etc.) about the fact that  $p$ ,  $S$  *wishes it not to be the case that  $p$* .

It is worth mentioning another relevant feature of wishing. Not only is it possible to be ambivalent in what one wishes, i.e., concurrently to wish it to be the case that  $p$  and to wish it not to be the case that  $p$ . More interesting, wishes are rationally blind to one another. That is, if I recognize that getting a certain wish of mine will exact an exorbitant cost in terms of other wishes of mine, this is no reason to give up or even to weaken that wish. For I may wish that getting that wish *did not* exact so high a cost; indeed, I may wish the world to be such that fulfillment of *all* of my present wishes would be mutually compossible. (R. S. Peters quotes a colleague who wished to be monogamously married to several women at once.)<sup>7</sup> This feature of wishing makes it possible for someone without irrationality to be angry at not having been appointed committee chairman, even though he strongly prefers not to have been appointed. It also makes it possible, again without irrationality, to have "mixed emotions" about something, i.e., to suffer concurrently a negative emotion and a positive emotion about the same fact: e.g., to be angry about not having been appointed committee chairman and at the same time to be very glad, even delighted, that one was not appointed committee chairman.

When  $S$  is angry about the fact that  $p$ ,  $S$  believes

that  $p$  and wishes it not to be the case that  $p$ . Since what  $S$  believes is the contradictory or at least the contrary of what  $S$  wishes, we can say that  $S$ 's wish is *frustrated*. (Since reference to "the fact that  $p$ " presupposes that  $S$ 's belief is true, the wish is frustrated in reality as well as "in thought"; but at this stage I am concerned only with frustration "in thought.") What is causing  $S$  to be angry is the belief and the wish in combination; it is, in other words, the wish-frustration that is causing  $S$  to be angry. *Anger always arises from wish-frustration*, if it is anger about a fact; or, granting the primacy of facts as urged in Sect. II, if it is anger about anything. The same holds for all the negative emotions we have considered.

We might try to fill out the analysis by adding to the Belief Requirement (BR) a parallel Wish Requirement (WR) as follows:

- (WR) If  $S$  is angry about the fact that  $p$ , then  $S$ 's wishing it not to be the case that  $p$  is sufficient for  $S$  to be angry, given some existing conditions that are not themselves sufficient for  $S$  to be angry.

Unfortunately WR, though true, does not give us what we want. For the conjunction of BR with WR does not insure that  $S$ 's believing that  $p$  works, so to speak, in conjunction with  $S$ 's wishing it not to be the case that  $p$ , to make  $S$  angry: i.e., it does not insure that the belief and the wish are parts of the *same* sufficient condition. We can easily repair the gap, however, by replacing BR and WR with the following single requirement, which shall be called the Belief-Wish Requirement (BWR):

- (BWR) If  $S$  is angry about the fact that  $p$ , then  $S$  believes that  $p$  and wishes it not to be the case that  $p$ ; and the conjunction of (1)  $S$ 's believing that  $p$  (2)  $S$ 's wishing it not to be the case that  $p$ , and (3) some other existing conditions is sufficient for  $S$  to be angry, where no two of these three conjuncts are jointly sufficient for  $S$  to be angry.

BWR holds, *mutatis mutandis*, for all the "negative" states listed at the beginning of this section. For the "positive" states, it holds if the "not" is deleted after "wishing it" (on each occurrence). For it is the *satisfaction* of the wish, not its frustration, that causes one to be pleased or delighted, at least

<sup>7</sup> See R. S. Peters, "Emotions and the Category of Passivity," *Proceedings of the Aristotelian Society*, vol. 62 (1961-62), esp. pp. 126-129. Peters rightly stresses the differences between wishing and wanting. He does not speak of the wish as playing a causal role, as I do.

in those cases where one is pleased or delighted about a fact. In referring to BWR henceforth I shall understand "angry" to be replaced by the predicate under discussion; and I shall understand the "not" to be deleted, where one of the "positive" predicates is under discussion.

The BWR principle helps us to decide what *S* is angry about in cases of the following sort. Suppose that Sam is angry about ( $F_1$ ) the fact that the fellow in the neighboring apartment plays the violin every evening. Among the *grounds* on the basis of which Sam is angry about  $F_1$  are ( $F_2$ ) the fact that the walls are so thin that any loud activity next door prevents Sam from sleeping, and ( $F_3$ ), the fact that Sam has to go to work every day at 6 A.M. (He is angry about  $F_1$ , it would be natural to say, "because of"  $F_2$  and  $F_3$ .) Now it might be held—correctly, I think—that in such a case Sam has three beliefs (corresponding to  $F_1$ ,  $F_2$ , and  $F_3$ , respectively) each of which satisfies the earlier requirement BR. Given only BR, Sam might be angry about  $F_1$  or about  $F_2$  or about  $F_3$ ; or he might be independently angry about any two or even all three of these facts, just as in the earlier example Daddy was independently angry about his car's being laid up and about Junior's having ruined the Hongkong shots. We need the stronger requirement BWR in order to decide which of these facts, if any, Sam is angry about. BWR requires that we ask, e.g., "Does Sam wish not to have to go to work at 6 A.M.? And is this wish sufficient for him to be angry, when conjoined with his belief that he has to go to work at 6 A.M., etc.?"

BWR gives us the basic analysis of being  $\phi$  about a fact, where ' $\phi$ ' is replaced by "angry" or by any other of the words listed at the head of this section or of Sect. I. I claim not only that BWR holds without exception but also that, aside from three complications to be noted below, it gives us all that we need to decide what fact or facts a person is angry (etc.) about: i.e., it gives us a complete set of truth-conditions for sentences of the form "*S* is angry (etc.) about the fact that *p*."

*First complication.* Suppose that Sam believes that he has to go to work every day at 6 A.M. and wishes not to have to go to work every day at 6 A.M. Suppose further that, because Sam does not have his wish in this matter, his wife is upset; and because she is upset, she cannot pay proper attention to her driving. Finally, her inattentive driving causes Sam

to be angry. It could be argued that in such a case Sam's wish-frustration over having to go to work at six is, given the intermediate linkage described, sufficient for Sam to be angry. Yet he is not angry about anything more than the fact that his wife is not paying proper attention to her driving.

What is required is a formula that will filter out all such counter-examples yet admit intermediate linkages of the right sorts. (Similar formulas are required by the belief-want analysis of intentional acts and the causal theory of perception.) But the problem of constructing a formula that will admit all and only the right sorts of intermediate linkages does not seem to me at all crucial to the acceptability of the basic analysis. Once we have narrowed the gap in our analysis to this problem of intermediate linkages, the analysis can no longer be fundamentally threatened by exotic counter-examples. They become tools, rather, for hacking out the needed formula.

*Second complication.* Some emotion-types, such as anger and indignation, are set off from others by their relatively specialized causal structures. If *S* is angry about the fact that *p*, then *S* believes not only that *p* but also that some agent—in standard cases a person or persons—by act or omission helped to bring it about that *p*.<sup>8</sup> (In some cases '*p*' is itself a description of an act, e.g., where '*p*' = "Sam's neighbor plays the violin every evening.") The causal structure for *indignation* includes, in addition, a belief that it is (putting it roughly) *unjust* that *p*. What is more, these additional beliefs must stand in certain relations to the elements of the basic causal structure described in BWR. For example, if Sam is *indignant* that his neighbor plays the violin every evening, then Sam wishes him not to. But this "wish" of Sam's must not be based entirely on considerations of interest, whether his own or another's; it must also be based on his belief that it is unjust that his neighbor plays the violin every evening. If moral beliefs play no role in shaping his attitude, Sam is not *indignant* but at most angry or resentful about the fact that his neighbor plays the violin every evening.

On the other hand, for many of the words listed at the head of this section, the basic analysis of being  $\phi$  about a fact is adequate without specialized amendments of the sort required, e.g., for "angry" and "indignant." Among these are "glad," "sad," "unhappy," and "upset." And for those that do

<sup>8</sup> There may be peripheral cases in which nothing that is properly an agent is believed to have brought it about that *p*. It is raining as I write this, and a friend of mine claims to be angry about the fact that it is raining. He doesn't believe in Jupiter Pluvius, but I am inclined to think that at the moment he must see "the world" as plotting against him.



require specialized structures, BWR provides at least the framework for further analysis.

*Third complication.* Any sentence of the form "*S* is angry about the fact that *p*" presupposes the corresponding sentence '*p*'. Thus it might not be true that *S* is upset about the fact that *p*, even though the Belief-Wish Requirement is satisfied: for it might not be true that *p*. But there is more to this complication than meets the eye, as I shall try to show in the section that follows.

## V. KNOWLEDGE: THE MISSING LINK

It is common to suppose that when we say of someone that he is, e.g., angry about this or that, we are saying something, essentially, just of him. One extreme form this supposition takes is the doctrine, criticized in Sect. III, that what a person is angry about is just a matter of how he feels. If it so happens that the form of words we use to describe his state carries a presupposition that what he is angry about is true or is a fact, this is incidental.

Yet there seems to be no form of words that entails everything that is entailed by sentences of the form "*S* is angry about the fact that *p*" except that it does not entail (or presuppose) '*p*'. Neither English nor any of the (few) European languages I have looked into has such a form. (To avoid the presupposition of '*p*' we usually revert to sentences of the form "*S* is angry because *S* believes that *p*": e.g., "Jules is angry because he believes that Jim has phoned Jane." In context this sentence may be used to assert what is asserted in "Jules is angry about the fact that Jim has phoned Jane," but this is not solely a function of sentence-entailments.)

We can show the presupposition of '*p*' to be something more than an incidental matter—a linguistic surd or brute fact, so to speak—if we can show that when someone is angry about the fact that *p*, the causal structure underlying his anger includes not only the elements specified in BWR but also the event or condition referred to by '*p*' (gerundized). To show that this is indeed the case, I must appeal to an earlier paper<sup>9</sup> in which I have sought to establish that if *S* is, e.g., angry or

annoyed or pleased that *p*, then *S* not only believes but *knows* that *p*. The arguments are applicable to any of the emotion words with which the present analysis is concerned, when followed by any of the factive complements indicated in Sect. I. In the discussion that follows I shall assume as a further "requirement," beyond the Belief-Wish Requirement (BWR), the following Knowledge Requirement (KR):

(KR) If *S* is angry (etc.) about the fact that *p*, then *S* knows that *p*.

KR together with BWR constitute what we might call the amended basic analysis.

KR is important to my present purpose for the following reason. Current in the literature are several causal analyses of "*S* knows that *p*."<sup>10</sup> These differ from one another in various details. But all, or almost all, are agreed on one point: viz., that if *S* knows that *p* then *S* would not have believed that *p* if it were not the case that *p*. Not all epistemologists are agreed even on this point, needless to say; but analyses that hold at least this much have won wide acceptance. And the use to which I am able to put this point can only strengthen its credibility.

Given this common denominator of most causal analyses of "*S* knows that *p*," and given my contention that when *S* is angry about the fact that *p*, *S* knows that *p*, we can see that the event or condition referred to by '*p*' is indeed a part of the causal structure underlying *S*'s anger, when *S* is angry about the fact that *p*: to be more specific, it is causally necessary, under the circumstances, for one of the elements in the causal structure delineated in BWR, namely *S*'s belief that *p*. This means that, aside from cases in which *S* is independently angry about several matters, the event or condition referred to by '*p*' is causally necessary, under the circumstances, for *S* to be angry.<sup>11</sup> For example, prior to his discovery that Junior had ruined the Hongkong shots, Daddy would not have been angry had it not been the case that his car would be laid up at the shop. It can also be shown that the event or condition referred to by '*p*' is causally sufficient for *S* to be angry, even where *S* is independently angry about several matters at once;

<sup>9</sup> "Emotions and Knowledge," *The Journal of Philosophy*, vol. 66 (1969), pp. 408–413. See also Peter Unger, "The Wages of Skepticism," *American Philosophical Quarterly*, vol. 10 (1973), pp. 177–187.

<sup>10</sup> See Alvin I. Goldman, "A Causal Theory of Knowing," pp. 357–372, and Brian Skyrms, "The Explication of '*X* knows that *p*,'" both in *The Journal of Philosophy*, vol. 64 (1967); Peter Unger, "An Analysis of Factual Knowledge," *The Journal of Philosophy*, vol. 65 (1968), pp. 157–170; and Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy*, vol. 49 (1971), pp. 1–22.

<sup>11</sup> I assume that "*x* is necessary for *y*" and "*x* is sufficient for *y*" are each transitive relations, even if "*x* is a cause of *y*" and "*x* causes *y*" are not.

though with some complications, especially where '*p*' is future-dated.

Thus it is a mistake to suppose that to say of someone that he is angry about this or that is just to say something of him, with at most an incidental implication that what he is angry about happens to be the case. We are also saying something about "external" events or states of affairs, in asserting that certain causal relations obtain between them and certain states of the subject. This conclusion squares with common sense and ordinary language. Common sense says that when you are angry about, e.g., Jim's having phoned Jane, Jim's having phoned Jane *affects* you in some way: viz., *it makes you angry*. You are angry *because* Jim has phoned Jane. In this matter alone my analysis of the relation between anger and what it is about has two advantages over Kenny's: it is consistent with the systematically causal locutions used to express that relation and hence obviates the need to explain away their causal implications; and, more impressive than this negative virtue, it actually explains why we use such locutions to express that relation.

To summarize briefly, the amended basic analysis of being angry (etc.) about the fact that *p* consists in the Belief-Wish Requirement (BWR) conjoined with the Knowledge Requirement (KR). What BWR says, in effect, is that to specify what *S* is angry (pleased) about is to specify a particular wish-frustration (wish-satisfaction) that is causally sufficient, in the circumstances, for *S* to be angry (pleased). But BWR is concerned only with satisfaction or frustration "in thought," i.e., with the congruence or incongruence of wish and *belief*; whereas in this section, using KR, we have shown this merely to be an intermediate link between *S*'s being angry (pleased) and satisfaction or frustration "in reality," i.e., the congruence or incongruence of wish and *fact*.

## VI. SOME CONSEQUENCES OF ACCEPTING THE ANALYSIS

1. It is sometimes suggested that the central or standard cases by which the concept of anger is to be understood are cases in which a person is, not just angry, but angry *about* something. Thus it is not just a contingent fact that anger is, at least in standard cases, anger *about* something. Even the stronger suggestion may be made that it is conceptually impossible that someone be angry yet not angry about something. If my analysis is assumed, then either suggestion entails that it is not a con-

tingent matter that anger arises from certain causes rather than others: viz., that when someone is angry, he has (at least in standard cases) a wish-frustration that stands in a certain causal relation to his anger, as specified in BWR. The stronger suggestion, that it is conceptually impossible that someone should ever be angry yet not angry about something, has on my analysis the following interesting consequence: that there can be no such thing as an anger agent. Nothing, neither a drug nor anything else, can induce anger without the help of some wish-frustration, as specified in BWR. This, of course, would make the experiment described in Sect. III a conceptual impossibility.

Whatever we say in the case of anger, it is clear that one is never *indignant* without being indignant about something. This is not to say that one can at the time always *state* what one is indignant about, beyond saying "some injustice or other." The point can be expressed in terms of the proposed analysis, as follows. What may symptomatically appear to be indignation actually is not, if it is not generated by the right causal structure, including an agency-belief and (roughly) an injustice-belief; otherwise what we have is at most anger or resentment, or perhaps mere displeasure with some state of affairs. Thus there can be no drug that acts as an *indignation agent*. I am not quite so sure, however, what to say about a state that symptomatically resembles *anger* but is not produced by an agent-caused wish frustration. The aboutness of anger seems to me to be a very important semantic feature, for without it anger would be little more than a strange psychosis that strikes people inexplicably from time to time, saddling them with ugly desires and bizarre physiological disturbances. Because symptomatic anger that is not about something is so rare, however, we do not have a contrasting term that enables us to say, "This is not true anger but at most —." Were such cases to become common, I should think they would soon acquire a name of their own—most likely a name that clearly marks the state as a disease. Then we should be able to say without reservation that no one is *angry* unless he is angry about something.

2. Just as, in general, there are alternative true descriptions under which someone's act is intentional—"He flipped the switch," "He turned on the light," and so on—so there generally are alternative true specifications of what someone is angry about. To return to an early example, even supposing that Sam is not *independently* angry about several things at once, we might still say that

he is angry about "the fact that his neighbor plays the violin every evening," "the fact that he is prevented from sleeping every evening," "the fact that he is unready for work at six in the morning," and so on. One of the virtues of the present analysis is that it permits us to explain this. For what we have in this case are three frustrated wishes: viz., that the neighbor not play the violin every evening, that Sam not be prevented from sleeping, and that Sam not be unready for work. These wishes, presumably, all derive from one "basic" wish, perhaps the last of those mentioned or perhaps some further wish of Sam's, e.g., not to risk being fired from his job. That is why I say that, although Sam is angry about several facts at once, he is not *independently* angry about several facts at once. It is plain that there are many parallels to be explored here between my analysis of being angry (etc.) about something and the belief-want analysis of intentional action.

3. The proposed analysis does not conflict with the claim that emotion sentences satisfy certain criteria of *intentionality* designed to mark off mental states from physical. It does, however, have the interesting consequence that the intentionality of emotion sentences is derivative from the intentionality of sentences that ascribe beliefs and wishes. For example, the sentence, "Smith is angry about his being the last person invited to the party," is intentional on one of the three criteria (viz., non-substitutivity) originally proposed by Chisholm.<sup>12</sup> But this sentence is true just in case a certain sentence of the following form is true: "Smith is  $\phi$  because he believes that he is the last person invited to the party and wishes not to be the last person invited to the party." But any sentence of this form would be intentional on the non-substitutivity criterion, even if ' $\phi$ ' were replaced by some blatantly physical predicate; for the non-substitutivity is introduced entirely by the belief and wish clauses. Thus, even if the criterion were successful in distinguishing mental states from physical, it would not show that *being angry* is *per se* a mental and not a physical state. This conclusion is vitiated, however, if we grant that as a matter of definition to be angry is to be angry about something. For in that case, it may be argued, "Smith is angry" is true just in

case *some* intentional sentence of the form "Smith is angry about  $X$ " is true; hence Smith is angry just in case he is in a mental, non-physical state described by at least one of these intentional sentences. In any case, if being angry is a disposition to acquire *desires*, etc., the advocate of intentionality is back in business: for now we should have to cope with the intentionality of sentences about desires.

4. According to the analysis here proposed, our presumed ability to discover in many cases what an individual is angry about is an ability to secure at least implicit knowledge of the causal structure underlying his anger. Coupled with our ability to recognize that this or that act or physiological condition is a manifestation of anger, we can know at least implicitly that the act or condition was generated by a certain causal structure. What the proposed analysis does is to convert our implicit knowledge of this causal structure into explicit knowledge. Thus in a way it adds to our explicit knowledge of the causes of particular acts and psychogenic physiological conditions. Viewed in this way, the analysis, coupled with the presumption that we can, *from behavioral and contextual information alone*, recognize what someone is angry about, is open to empirical testing. Neurophysiological studies should reveal the appropriate causal chains between the various manifestations of anger and events or conditions at those functional centers of the brain, presumably within the cerebral cortex, that are associated with believing and wishing. (Where  $S$  is angry about a *fact*, e.g., that Snerd was appointed office manager, a chain should be traceable ultimately to the appropriate environmental event or condition, viz., Snerd's being appointed office manager: for in such a case  $S$  not only believes but *knows* that Snerd was appointed office manager.) If it can be shown that these chains do not exist, then we must abandon either the proposed analysis or the presumption that we are able to recognize, from behavioral and contextual information alone, what a person is angry about. It is highly unlikely that we would opt for the second alternative; it is therefore the analysis that would be on the line.<sup>13</sup>

University of Missouri, St. Louis

Received January 2, 1973

<sup>12</sup> Roderick, M. Chisholm, *Perceiving: A Philosophical Study* (Ithaca, N.Y., 1957), Ch. 11.

<sup>13</sup> This paper has benefitted, on one or another stage in its fitful progress, from the good advice of Fred Dretske, Peter Unger, Arthur Danto, Robert Audi, and several of my present colleagues.

## IV. EVALUATIVE ASYMMETRY

A. F. MACKAY

### I. THE DATA

- A. I favor taxing excess profits.
- B. Why?
  - A. If wages, prices, and rents are controlled, it is unfair to exempt profits.
  - B. But have you considered the consequences? The effect of levying an excess profits tax will be to discourage firms from maximizing profits, and the effect of that will be to destroy the profit motive as an incentive for efficiency, innovation, and risk taking.
  - A. But that's not a reason against imposing an excess profits tax.
  - B. Why not? Fairness is a point in favor of such a tax, but surely, having these bad consequences is a point against it. Of course, you might deny that it *would* have these consequences, or, for that matter, that they are all *that* bad. But if it will and they are, that makes an excess profits tax worse than it would be otherwise.
  - A. Not at all. You are assuming that to the extent that an otherwise good thing (excess profits tax) has bad consequences (inefficiency, economic decay), it is less good. But by the same token, to the extent that an otherwise bad thing (inefficiency, economic decay) has a good cause (excess profits tax), it is less bad. Thus these equal and opposite considerations cancel each other out. Whatever tendency the badness of inefficiency has to diminish the goodness of its cause is precisely balanced off and cancelled out by the tendency that the goodness of an excess profits tax has to diminish the badness of its consequences.
  - B. I see your point, but you are mistaken. No such symmetry exists.

### II. THE THEORY

Common sense holds that in assessing the value of a thing, it is relevant to consider the value (if

any) of its consequences. But is it ever relevant to consider the value of its cause? One answer to this is

- (1) The value of a set of consequences is always relevant to the value of what causes them, but the value of a cause is never relevant to the value of what it causes.

This claim is typically received with scepticism, a quick, unsuccessful search for a counterexample, and a continuing conviction that counterexamples exist, even if no satisfactory one can be turned up in a few moments thought. No doubt this is due to the universality of the claim—the occurrence of “always” and “never” rather than the more cautious “generally” and “rarely”—plus the suspicion that abstract value theory is, if not the home, at least among the last known addresses of hasty generalization. Still, I think that (1) is both true and important.

The following are variant formulations (only with a bit less generality) of (1):

- (1a) Having good consequences is a good-making feature of an object (i.e., a feature which an object is better for having), but having a good cause is not a good-making feature.
- (1b) Having bad consequences is a bad-making feature, but having a bad cause is not.
- (1c) A thing which has bad consequences is thereby less good than it would be otherwise, but a thing which has a bad cause is not thereby less good than it would be otherwise.
- (1d) A thing which has good consequences is thereby less bad than it would be otherwise, but a thing which has a good cause is not thereby less bad than it would be otherwise.

These indicate how the concept of *relevance* is to be taken. When it is said that the value of the consequence is *relevant* to the value of the cause, but not *vice versa*, it is meant that the value of the con-

sequence tends to affect or modify (i.e., increase or decrease, as the case may be) the value of the cause, but not *vice versa*.

The following is a typical situation to which (1) and its variants apply.

The recent discovery that widespread use of the pesticide DDT has among its consequences various ecological horrors, led many people to revise (downward) their assessment of the value of DDT. It did not lead people to revise (upward) their assessment of the value of the extinction of species. People said, "Since DDT has these disastrous consequences, it can't be such a good thing as we had previously thought." They did not say, "Since the silent Spring has such a good cause, it can't be as bad a thing as we had previously thought."

I shall refer to (1) and its variants as the Asymmetry Thesis. I mean thereby to contrast them with the Symmetry Thesis, which can take any of these forms:

- (2) The value of a cause is sometimes (always) relevant to the value of what it causes.
- (2a, b) Having a good (bad) cause is a good- (bad-) making feature of an object.
- (2c, d) A thing which has a bad (good) cause is thereby less good (bad) than it would be otherwise.<sup>1</sup>

Attempts to find a counterexample to the Asymmetry Thesis thus constitute attempts to describe a situation in which the Symmetry Thesis obtains. Many such attempts fail to observe the following distinction. It often happens that a thing will have such a good cause that we are willing to accept its (on balance less) bad consequences. The value of the cause, we say, overrides or outweighs the disvalue of its consequences.<sup>2</sup> This, of course, does nothing to vindicate the Symmetry Thesis which, as my interpretation of the concept of *relevance* shows, requires that the valuable cause (directly) make the disvaluable consequences less bad. When the value of the cause overrides or outweighs the disvalue of the consequences, it does not thereby

make them less bad. It induces us to accept them, to tolerate them, to put up with them (being more than adequately compensated for the sacrifice), but it doesn't make them any better. The difference here is between something's compensating you for incurring a burden, and something's lessening the burden.

The general point can be put another way. Suppose some (mildly) pernicious side effects are produced by two different drugs, one of which is absolutely indispensable to the treatment of a horrible disease, the other is marginally better than a placebo in the treatment of a non-serious cosmetic annoyance. Then however bad these side effects are, they are going to be just that bad when produced by either drug. They aren't worse when produced by the trivial drug, and better when produced by the indispensable one. Although, of course, we are more willing to put up with them when they are produced by the indispensable drug (when we need it). So, to describe a situation in which the value (positive or negative) of a cause *overbalances* the value (negative or positive) of its consequences is not to confirm the Symmetry Thesis, and hence, not to find a counterexample to (1).

The distinction between a thing's being made better or worse, and our being made more or less willing to accept or incur it, not only disarms many putative counterexamples to the Asymmetry Thesis, it illuminates our understanding of ordinary consequential evaluation as well. When (1) says that the value of consequences is relevant to the value of what causes them (but not *vice versa*), it means this: bad consequences reduce the value of their cause, good consequences increase it. Symmetry would be vindicated if this were not the case. The dialectical situation is something like this. Asymmetry involves the claim that (a) while value of consequences do affect value of causes, (b) value of causes do not affect value of consequences. To establish Symmetry, one might deny (b) while admitting (a)—a move often fostered by the confusion discussed above—or, equally, deny (a) while admitting (b). That is, with regard to whether or not evaluative influence is exerted, one might say either,

<sup>1</sup> Each variant of the Symmetry Thesis affirms what the corresponding variant of the Asymmetry Thesis denies but does not deny what its counterpart affirms. For instance, (2) does not deny that the value of consequences is relevant to the value of what causes them. Also, the Symmetry Thesis comes in a stronger and a weaker form. If we think of the contrast between (1) and (2) as being that (1) claims something never is the case which (2) claims is the case, (2) may contradict (1) by claiming that it always is the case (strong form), or that it sometimes is (weak form).

<sup>2</sup> Roderick Chisholm calls this phenomenon "balancing off," although he restricts his discussion exclusively to intrinsic value. See his "The Defeat of Good and Evil," *Proceedings and Addresses of the American Philosophical Association*, vol. 42 (1968-69), pp. 21-38.

- (a') Consequences do, but so do causes, or
- (b') Causes don't, but neither do consequences.

We must consider then, the following sort of objection.

When we say that a thing is less bad for having good consequences than it would be otherwise, we mean merely that the value of the consequences (at least partially) compensates us for the disvalue of the cause, thereby making us more willing to accept it than we otherwise would be. What we don't mean is that the value of the consequences directly affects the value of their cause, reducing or increasing it, as the case may be. No, the cause has whatever value it has, regardless of its consequences. A little attention to the distinction between intrinsic and extrinsic value would prevent a great deal of confusion here.

For instance, take the DDT case. The use of DDT, considered by itself, apart from its consequences, is evaluatively neutral. We only value DDT derivatively for its good consequences, chiefly, the elimination of malaria. So, strictly speaking, it never was DDT that we valued, it was the elimination of malaria. Therefore, when it was discovered that DDT caused ecological horrors, it wasn't that what had itself been good (DDT) became less good or maybe even bad, rather; it was that we were no longer willing to pay the price—which had been discovered to be enormously higher than previously thought—required to achieve the valued thing (elimination of malaria) using those means (DDT). The penalties of using DDT have been discovered to outweigh the benefits, and we incur, so to say, negative compensation. But the value of the use of DDT itself hasn't changed. It has whatever value it has, regardless of the value of its consequences.

This objection, however tempting, is fundamentally mistaken. At bottom it involves the view that

extrinsic value really isn't value at all. I shall not here delve into all the varieties of goodness, but I shall insist that they all are, after all, varieties of goodness. Thus, that *X* has consequential value cannot be explained to mean nothing more than that *X* causes, leads to, or, is a means for achieving *Y* (which is good). To say that *X* is good because it causes *Y* (no matter what we go on to add like, e.g., "and it wouldn't be any good at all if it didn't") is to say that *X* is good. Whereas, to say merely that *X* causes *Y*—even if we go on to add that *Y* is good—is not to say that *X* is good. It is not to evaluate *X* positively, since it is not to evaluate *X* at all. Whether or not something causes something is a purely factual, descriptive feature of the thing, even if what it causes is good. To get an evaluation, along these lines, we must add that a thing is good for causing a good thing.

More generally, if *X* is extrinsically valuable, *X* is valuable. Thus, the tendency some people have to denigrate extrinsic value—"if it's only extrinsically valuable, it isn't really valuable" or "it never was DDT that we valued, it was the elimination of malaria"—is quite unwarranted. From the fact that *X* is only valuable because of *Y*, it does not follow that *X* is not valuable: quite the reverse.

It is useful in this connection to think of the distinction between intrinsic and extrinsic value in quasi-modal terms. The leading idea is this: A thing is intrinsically valuable if it is valuable in every possible world in which it exists.<sup>3</sup> A thing is extrinsically valuable, on the other hand, if it is valuable in this world, and there is some possible world in which it exists but is not valuable.<sup>4</sup> So, a claim that something is intrinsically valuable is analogous to a claim of necessity, and a claim that something is extrinsically valuable is analogous to a claim of contingency. Hence, the denial that a thing which is (only) extrinsically valuable is really valuable, is like the denial that a contingent truth is really true.<sup>5</sup>

To return to the objection, it is true that when we say that DDT is less valuable for causing ecological

<sup>3</sup> *Ibid.*, Chisholm, p. 34.

<sup>4</sup> G. E. Moore's definition of intrinsic value in *Ethics*, and the famous method of isolation of *Principia Ethica*, can be thought of as establishing that a thing is intrinsically valuable (in this quasi modal sense) by requiring that it be valuable in the extreme (least favorable) case. Thus, if a thing would be valuable even in a world in which it existed "quite alone, without any further accompaniments or effects whatever,"—these being, presumably, the most adverse value conferring circumstances to be encountered—then it would be valuable in all possible worlds in which it existed. Claims of extrinsic value, on the other hand, are typically followed by "but only because" qualifiers. For example, *X* is good, but only because it has consequences *Y* which are good, or, but only because it contributes to the value of *Z*, a good whole. Thus, if *X* didn't cause *Y* or contribute to the value of *Z*, it wouldn't be good. Hence, in a possible world in which *X* existed, but *Y* (or *Z*) didn't, *X* wouldn't be valuable.

<sup>5</sup> Furthermore, the conclusion of the regress argument, that there couldn't be extrinsic value if there weren't intrinsic value, is like the claim that there couldn't be contingent truth if there weren't necessary truth. (Continued on next page.)

horrors than it would be otherwise, we don't mean that its intrinsic value is diminished. Intrinsic value is, by definition, independent of extrinsic value. What we do mean is that its extrinsic value is diminished or reduced. And since extrinsic value is value, we do mean that its value is diminished.

Suppose we have a thing  $X$ , its cause  $Y$ , and its consequence  $Z$ . We can ask, subjunctively, how would the value of  $X$  be affected were the value of  $Y$  to be different, and how, were the value of  $Z$  to be different? The answer, in the case of the cause, is that the value of  $X$  would not be affected at all. In the case of the consequence, the answer is that the value of  $X$  would be increased or diminished according as the value of  $Z$  increased or decreased.

The situation with bribes and threats is instructive here. When Jones offers Smith a bribe to throw the game, he is trying to make that course of action more valuable to Smith, thereby, presumably, altering its relative standing among all of Smith's options, and rendering it more eligible for choice. When Brown, on the other hand, threatens Smith to get him to throw the game, he also has in view altering Smith's relative preference rankings in such a way as to render throwing the game more eligible for choice. But he does this not by elevating (making more valuable) the option of throwing the game, but by depressing Smith's other options. This is the way it is with changes in the value of the consequences *vis a vis* changes in the value of the cause. Changes in the value of the consequences work by way of changing the value of the thing, all else remaining the same. Changes in the value of the cause alter relative preference rankings (when they do) by way of changing the value of other things, the value of the thing remaining the same.

To sum up then, the fact that a bad thing has an overwhelming good cause might indeed make us more willing to accept it than we would be otherwise. As likewise, the fact that a bad thing has overwhelmingly good consequences will make us more willing to accept it than we would have been otherwise. However, the latter, as opposed to the former, does this by way of making the thing better than it would have been otherwise. In particular, the fact that the widespread use of DDT has as a

consequence various ecological horrors makes it worse than it would be otherwise. And, should it turn out 20 years hence that the widespread use of DDT has hitherto unsuspected good consequences—say, it not only eliminates malaria, but ignorance, poverty, and war as well—that would make it better than it would be otherwise. But no matter how good it ultimately turned out to be, that wouldn't make the extinction of species one whit the better by virtue of having such a good cause.

Another confusion, which is sometimes a source of alleged counterexamples to (1), is this. It is sometimes thought that judgments of extrinsic value, while they do attribute value to something, nevertheless don't attribute it to what they appear to, nor in the way they appear to. Consider the judgment, " $X$  is good because it causes  $Y$ ." This appears to say that  $X$  is good, that it is  $X$  (not something else) which is good, and furthermore, it does not appear to be comparing the values of  $X$ ,  $Y$ , or the complex object ( $X + Y$ ) with each other or with anything else. The view currently under discussion denies this. It holds that a proper analysis of judgments of extrinsic value shows them to be quite different from what they appear. The view runs something like this.

When someone says that  $X$  is less good for causing  $Y$  than it would be otherwise, what he is really saying is that the complex object ( $X + Y$ ) is less good than the simple object  $X$  alone. Thus, all judgments of extrinsic value are really comparative judgments of the value of complex objects. They are all applications of Brentano's Axiom:

When  $A$  is something of positive value and  $B$  is something of negative value, then  $A$  is more valuable than the conjunction of  $A$  with  $B$ , and this conjunction is more valuable than  $B$ .<sup>6</sup>

This, of course, is perfectly symmetrical with respect to causes and consequences. Exactly as (a) the complex object (good cause + bad consequence) is less valuable than the simple object (good cause) alone, so (b) that complex object is more valuable than the simple object (bad consequence) alone. Since the former is

---

There are enough dis-analogies between these two sets of distinctions that the characterization "quasi-modal" is in order. Among these is that it is possible for a thing to be both intrinsically good and extrinsically bad, but not for a thing to be both necessarily true and contingently false. Herein may lie a motivation for treating intrinsic and extrinsic value as different kinds of value.

<sup>6</sup> As formulated by Nicholas Rescher in *An Introduction to Value Theory* (Englewood Cliffs, 1969), p. 58.

what is meant by saying that the value of the consequence is relevant to the value of the cause, the latter is what is meant by saying that the value of the cause is relevant to the value of the consequence. Thus, the complex object (elimination of malaria + extinction of species) is worse than the simple object (elimination of malaria) alone, but better than the simple object (extinction of species). Hence this same fact, when looked at one way, makes the use of DDT worse, and when looked at the other, makes the extinction of species better.

The crucial question for this objection is whether it embodies an acceptable analysis of consequential evaluations. It holds that when one says

- (a) Since eating scrambled eggs causes nuclear war, it is bad.

one really means

- (b) The complex object (eating scrambled eggs + nuclear war) is worse than the simple object (eating scrambled eggs) alone.

Now in the first place, (b) is true while (a) is false, hence the one can hardly be an analysis of the other. Secondly, the comparative evaluation, (b), yields a judgment that is peculiarly unfocused. The fact that the complex object (eating scrambled eggs + nuclear war) is considerably less valuable than the simple object (eating scrambled eggs) alone, does not reflect at all, much less adversely, upon the merits of eating scrambled eggs. What, we might ask, wouldn't form, in combination with nuclear war, a complex worse than itself? Yet, if (a) were to be true, if eating scrambled eggs did cause nuclear war, that would certainly reflect adversely upon the merits of eating scrambled eggs. In short, we are saying something considerably more incriminating about the use of DDT when we say that it causes the extinction of species than when we say that it, plus the extinction of species, together form a complex which is worse than itself alone. Hence, the latter does not give an acceptable analysis of the former. And it is precisely this, of course, that marks the asymmetry. Because in fact we are not saying anything more ameliorating about the extinction of species when

we say that it is caused by the use of DDT than when we say that the complex object (use of DDT + extinction of species) is better than the extinction of species alone. In neither case are we saying anything which bears upon the value of the extinction of species.

Thus I conclude this part of my defense of (1). I have not argued directly for it, since I consider it to be an axiom of evaluation, and one does not derive one's axioms from anything more basic. What I have attempted to do, borrowing strategy from Moore, is point out and argue against certain superficially tempting views which, were they to be true, would lead people to reject (1). These were (a) the view that extrinsic value isn't really value, (b) the view which fails to distinguish between the value of one thing's affecting the value of another, and the value of a thing's balancing off, or overbalancing, the value of another, and (c) the view which holds that judgments of extrinsic value are really judgments of the comparative value of complex objects. I suspect that any plausible putative counterexample to (1) will depend, for its plausibility, on one or more of these three mistakes.

I shall now offer a few brief arguments in favor of (1). In keeping with my view of its status as an axiom, these arguments will be non-constructive, deducing absurdities from the supposition that the Symmetry Thesis holds.

(A) *Theodicy*: Suppose the Symmetry Thesis true. Then the fact that a thing has a good cause makes the thing less bad than it would be otherwise. Hence we have a trivial solution to the Problem of Evil. There is and can be no Problem of Evil to be solved. The very statement of the Problem—What justification has God for causing evil?—presupposes that it is possible for God to cause evil. But that is not possible if the Symmetry Thesis is true. If a thing is less bad for having a good cause, and if the better the cause the less bad the thing caused, then nothing with an infinitely good cause can have any badness in it. God can cause no evil since, try as He might, if a thing is caused by Him, then by virtue of having such a good cause it cannot be bad.<sup>7</sup> But there is a Problem of Evil, and it is possible for God to cause evil, hence, the

<sup>7</sup> This is not an epistemological point. Speaking from a fallibilist epistemological stance, it might be argued that since God is so good, it is more likely that one is mistaken in his unfavorable appraisal than that God, who is so good, caused something so bad. In short, the remark, "If God did it, it can't be all that bad," is ambiguous. It might be an epistemological disclaimer meaning, roughly, "Since God is so good, He wouldn't do anything that bad. Hence, appearances to the contrary notwithstanding, I must be mistaken in thinking it that bad." Or, it might have the meaning which the Symmetry Thesis would underwrite, namely, "God's doing it makes it less bad than it would be otherwise."



Symmetry Thesis, which implies the opposite, is false.

(B) *Ethics*: Suppose the Symmetry Thesis true. Then any given consequence would be less bad if it had a better cause. Now suppose a good man learns of a plot by bad men to do something which would have very bad consequences (say, assassinate the archduke). Then setting aside all the various *prima facie* duties the good man has—to notify the police, warn the archduke, etc.—as being, due to some unfortunate material circumstance, not open to him to perform, he has the one remaining duty, underwritten by the Symmetry Thesis, namely, to intervene, preempt the plot, and assassinate the archduke himself. Why does he have this duty? Because by so doing he can make the consequences of the assassination less bad. Since, by hypothesis, he is a good man and the original plotters were bad men, if he does it the consequences will be due to a better cause than if they did it. And if due to a better cause, then less bad. But in fact, such a man would have no such (even *prima facie*) duty. Hence the Symmetry Thesis, which implies that he would, is false.

(B') *Theodicy Corollary*: Assume that evil's being caused by God makes it less bad than it would be otherwise. Then the question arises, not why God causes so much evil, but why He doesn't cause more? Not, of course, that He should increase the absolute amount of evil in the universe; just that when He foresees some impending evil about to be caused by natural causes, He should intervene, preempt the natural causes, and cause it Himself—presuming that He has some morally sufficient reason for not just preventing it altogether. The people should mutter and complain, "If only that tornado had been caused by God, it wouldn't have been so bad." Job's comforters should have said, "Cheer up. Things could be worse. Suppose these boils hadn't been caused by God." But this is absurd, and so the Symmetry Thesis is false.

### III. EXTENSION OF THE THEORY

Now I should like to extend the scope of these considerations to show that there is not just one,

<sup>8</sup> If this were the case, (1) would not be an axiom in one of my senses, because it would be deducible from some more basic truths about causality and agency. Even so, it would still be a conceptual truth, that is, one to be established or refuted on the basis of a *priori* reflection.

<sup>9</sup> This case raises many complex and difficult issues. For instance, it's very hard to get a coherent interpretation of what the symmetry case would be like if it were true. In particular, what is it exactly whose value would get modified in virtue of what? Also, quasi-modal considerations again obtrude, this time concerning essential and accidental properties. There doesn't seem to be any difficulty in asking what effect it would have on the value of a thing if its consequences had been different. But when we ask what effect it would have on the value of a thing if it were to be an instance of a different kind, some people get possible worlds vertigo.

but a number of these Asymmetry Axioms. This is of interest on several counts. First, so far as I know, this whole class of axioms has gone unnoticed in previous work on abstract value theory. Secondly, their consideration illuminates much concerning the less well known "kinds of value," e.g., contributory, inherent, etc. Thirdly, the fact that these asymmetries occur for types of evaluation other than consequential shows that the most natural explanation for the consequential asymmetry probably isn't correct. Initially, this latter asymmetry seems to be derivative in some way from the temporal asymmetry of causality, plus, perhaps, the alleged fact that as agents our primary orientation is toward the future.<sup>8</sup> The following extension seems to me to preclude any such explanation.

Many have held that consequential assessment is not the only sort of evaluatively relevant thinking. W. D. Ross, for instance, claims that the nature of a thing—its being of a certain sort, kind, or type—is also relevant to its assessment. Thus, if some contemplated act turns out to be an instance of theft, or infidelity, or murder, then that in itself, consequences aside, is evaluatively relevant data. As Ross might put it, in virtue of being an instance of a certain kind, an act might be *prima facie* right or wrong. We can easily adapt this to other than act appraisals. We might say, for example, that a state of affairs, in virtue of being of a certain kind—e.g., one in which the good are unhappy and the wicked happy—is *prima facie* bad, or, at least, that such a state of affairs is less good than it would be if it were of another kind—e.g., one in which the good are happy and the wicked unhappy.

But notice the asymmetry here. We think (or at least Ross does) that the kind of thing something is, is relevant to its value. We say (Ross does) that *X* can't be all that good considering it is an instance of kind *K*. But we don't think that since *X* (which is otherwise so good) is an instance of kind *K*, then being an instance of kind *K* can't be all that bad.<sup>9</sup> The Asymmetry Axiom which governs this sort of evaluation is this:

- (3) The fact that a thing is an instance of a certain kind can be relevant to the value of

the thing—"relevant" in the sense of making it better or worse than it would be otherwise—but not *vice versa*.

The fact that a given act is an instance of infidelity can make that act worse, but can't make infidelity better.

I shall consider the next group of cases from a different perspective, one which focuses attention on the significance of some terminology gaps. Consider the technical terminology used for classifying the various "kinds of value." We have a name for the kind of value a thing has in virtue of having good consequences (*extrinsic*, or, *consequential* value), but no name for the kind of value a thing has in virtue of having a good cause. We have a name for the kind of value a thing has in virtue of being instrumental toward securing, achieving, or promoting something (*utility* value), but no name for the kind of value a thing has in virtue of being securable, achievable, or promotable by means of a certain instrumentality.<sup>10</sup> We have a name for the kind of value a thing has in virtue of contributing to the value of a good whole (*contributory* value), but no name for the kind of value a thing has in virtue of being a whole whose value is contributed to by a certain thing. We have a name for the kind of value a thing has in virtue of its experiencing being good (*inherent* value), but no name for the kind of value an experience has in virtue of its being the experiencing of something which is good.

So we have terminology gaps. Our periodic table of kinds of value has some glaring omissions, systematically arranged. The question is, Why these gaps? Technical terminology is easy to invent. As a name for the opposite of *contributory* value, we might have had *recipient* value; for the opposite of *inherent* value, *exherent* value, and so forth. And we can do better than just coin terms, we can fit them out with the usual style of definitions. Here are a few standard definitions:<sup>11</sup>

- A: *Utility Values*: Things that are good because of their usefulness for some purpose.
- B: *Extrinsic Values*: Things that are good because they are means to what is good.
- C: *Inherent Values*: Things that are good be-

cause the experience of contemplating them is good or rewarding in itself.

For each of these we can construct symmetrically opposite, counterpart definitions:

- A': *Purposive Values*: Things that are good because they are purposes furthered by some useful thing.
- B': *Intrinsic Values*: Things that are good because they are ends to which good things are means.
- C': *Exherent Values*: Things which are good because they are the contemplative experiences of things which are good or rewarding in themselves.

It is not, in my opinion, a linguistic accident that we lack this extra bit of terminology. Each of these definitions (A-C, A'-C') embodies an implicit, substantive claim to the effect that such-and-such a characteristic is in fact a good-making feature which an object might have. In the case of the constructed definitions (A'-C') that implied claim is false: hence the terminology gaps.

The reason for the gaps then is this. For each different "real" kind of value there is a good-making feature such that if an object has that feature then it has that kind of value. Since, for example, it is a good-making feature to be the cause of a good thing, and since it is also a good-making feature to contribute to the value of a valuable whole, and since these are different good-making features, we create technical terminology to distinguish them. We say that when a thing has one of them it has *extrinsic* value, when it has the other, *contributory* value. Furthermore, if being the cause of a good thing were not a good-making feature, there would be no need for distinctive terminology to mark it in evaluative contexts, and in particular, it would be a mistake to invent a name for a kind of value which an object would have in virtue of having this characteristic—since, given that this would not be a good-making feature, no object would have any kind of value in virtue of having this feature. Such, I claim, is the case with the characteristic "being caused by a good thing." It is not a good-making feature, and

<sup>10</sup> What about feasibility? We think a plan or project better for being feasible, i.e., achievable by means of non-costly instrumentalities ready to hand. But the question is not whether feasibility is a good-making feature. The question is whether "feasibility value" would be the symmetrical opposite of utility value. Now utility value is the kind of value a thing has in virtue of being instrumental toward securing some goal. Its symmetrical opposite would be the kind of value a goal derives in virtue of being securable by a certain instrumentality. But feasibility is not a feature of a goal, it is a feature of plans or projects, which are a sort of instrument or tool for achieving goals. Hence, "feasibility value" would be a specialized sub-case of utility value.

<sup>11</sup> From William Frankena's *Ethics* (Englewood Cliffs, 1963), p. 66.

hence we have no special terminology—no “intrinsic value”—contrasting, but on a par, with extrinsic value. Likewise, “having its value contributed to by a (good?) part”—the constructed counterpart of the good-making feature associated with contributory value—is not a good-making feature, and hence the terminology gap: no “recipient value.” Also, “being the contemplative experiencing of something intrinsically good”—the constructed counterpart of the good-making feature associated with inherent value—is not a good-making feature, and hence we have no name for any kind of value which an experience might have in virtue of possessing this feature: no “exherent value.”

In short, it's not enough just to construct counterpart definitions and invent technical terminology. The substantive claim that such-and-such a characteristic really is a good-making feature has to be made out before any such definitions and terminology embodying this sort of claim will be acceptable. The burden of our Asymmetry Axioms is that many possible claims of this sort are false. (1), for instance, holds that while having good consequences is a good-making feature and having bad consequences is a bad-making feature, having a good cause is not a good-making feature, nor is having a bad cause a bad-making feature. Other Asymmetry Axioms are:

- (4) It is a good-making feature of a thing that it contributes to the value of a good whole, but it is not a good-making feature of a thing to have its value contributed to by a (good?) part.
- (5) It is a good-making feature of a thing to afford a good experience to some percipient, but it is not a good-making feature of an experience to be afforded to some percipient by a good thing.
- (6) It is a good-making feature of a thing to be useful for some purpose, but it is not a good-making feature of a thing to be a purpose furthered by some useful thing.

#### The questions

How many of these Asymmetry Axioms are there?

Is there any unifying rationale, any generating principle, for them?

What, in detail, is their content?

are questions to which I currently lack answers. But, as the preceding discussion has shown, they are closely related to the following question

What criteria must be met for a characteristic to count as a good-making feature of an object?

and this I should like to discuss.

Traditionally, something like this latter question would be answered by constructing a Theory of Value. One took an agreed upon list of good-making features and tried to find what they had in common in virtue of which they were good-making. For example, a hedonist logician might explain why simplicity is a good-making feature of proofs by showing that it is related in a certain way to pleasure. But such traditional accounts operate at a different level and answer a different question from the one now at issue. Traditional value theories try to justify a choice between such alternatives as *pleasure* and *self-fulfillment*, viewed as candidates for filling the blank in the context “being related in a certain way to —.” We, on the other hand, want justification for a choice between such alternatives as *having as a consequence* and *having as a cause*, viewed as candidates for filling the blank in the context “being related in way — to pleasure (or self-fulfillment, or whatever is the answer to the traditional question).”

The following is a schema for an answer to the question why a certain characteristic ( $\theta$ ) is a good-making feature:

$\theta$  is a good-making feature because it is related in way(s)  $\phi$  to  $\psi$ .

Traditional theories are primarily designed to provide rational grounds for instantiating the last variable ( $\psi$ ) in one way rather than another. We seek rational grounds for instantiating the penultimate variable ( $\phi$ ) in one way rather than another.

Even though we presently lack answers to these questions, we are presumably better off, as Moore would say, for having distinguished them, and for having clearly before our minds what questions are being asked, before attempting their answer.<sup>12</sup> At any rate, it seems clear that no general explanation of these evaluative asymmetries is forthcoming based on results concerning, say, causality, tem-

<sup>12</sup> The well known tendencies of intuitionist or objectivist theories to spawn epistemological and metaphysical difficulties is partly balanced off by the facility with which theoretically interesting questions can be formulated in such unreconstructed discourse. One need only attempt to pose, much less answer, questions concerning evaluative asymmetries, or the criteria for something's being a good-making feature, in speech acts theory or in “point of view” talk, to appreciate this point.

porality, and agency. These concepts apply to only one of the various asymmetries we have noted.

#### IV. APPLICATION OF THE THEORY

Finally, I should like to indicate one aspect of the importance of these considerations. Not only do they provide the basis for a refutation of classical psychological naturalism, they also refute any theory like that of Charles Stevenson in *Ethics and Language* which ties ethics too tightly to empirical psychology. The refutation proceeds as follows.

Any such theory, which isn't just incomplete at this point, must provide psychological analogues for our asymmetry axioms. Our evaluative asymmetries must be reflected in psychological asymmetries, or else there is something basic, true, and important about ethics which is not captured by the theory in question. For example, underlying (1) must be something like the following:

If a person has a positive attitude toward *A*, and a negative attitude toward *B*, and comes to believe that *A* causes *B*, the attitude toward *B* will modify the attitude toward *A*, but not *vice versa*. That is, he will come to have a less positive attitude toward *A*, not a less negative attitude toward *B*, as a result of discovering that *A* causes *B*.

The difficulty is, while our evaluative asymmetries are conceptual truths (non-contingent axioms), such psychological analogues will all be contingent empirical generalizations, very likely false. No theory which transforms conceptual truths into empirically hazardous contingent claims is acceptable. Thus they are all refuted.

Consider Stevenson's account in "The Emotive Conception of Ethics and Its Cognitive Implications":

Just how does this influence of thoughts upon attitudes take place? . . . Our approval of anything is strengthened or weakened depending on whether we approve or disapprove of its consequences. Suppose, for instance, that a man has conflicting attitudes towards *X*, and suppose that he later comes to believe that *X* causes *Y*. Now if he approves of *Y* . . . he will thereupon approve of *X* more strongly . . . The role of thought or cognitive inquiry in this example will be obvious: it establishes the ordinary causal proposition that *X* leads to *Y*. But we still have to explain why a belief of this proposition does anything more than satisfy a

scientific curiosity. Why does it strengthen the man's approval of *X*? One cannot easily hold, I think, that the belief has any power *in itself* to do this. It strengthens the man's approval of *X* only because *Y* too is an object of his approval. If *Y* were indifferent to him he would feel that any question about the relation of *X* to *Y* was foreign to his problem. His reasoning serves, then, purely as an *intermediary* between his attitudes: by connecting his thought of *X* with his thought of *Y* it also connects his attitude toward *X* with his attitude toward *Y*, letting the one be reinforced by the other.<sup>13</sup>

This is, of course, an updating of Hume's view of the place of reason in conduct determination. Reason alone (the mere discovery of truth and falsehood) can never provide an originating impulse to action. Only an attitude, or some other dynamic psychological phenomenon, can do that. Beliefs do, however, have a subsidiary role to play. They serve as conduits, pipelines, or transmission links which establish connections between our various otherwise disconnected attitudes and allow them to influence each other. Thus while reason is never an originating impulse to action, it can serve to channel, direct, and redirect pre-existing impulses.

This picture forces upon us the asymmetry question: Can the flow ever go in the other direction? Can the attitude toward the cause ever affect the attitude toward its consequence? If the answer is "Yes," the psychological theory is immediately refuted since it perforce must transform a conceptually true evaluative axiom into a false psychological generalization. So, presumably the answer will be "No." The attitude toward the consequence, then, can affect the attitude toward the cause, but never the reverse. Thus we have the psychological analogue of our first asymmetry axiom. But in contrast to the evaluative asymmetry claim, whether or not this psychological asymmetry claim is true is a matter of contingent fact, to be settled experimentally by psychologists, not by philosophers using the method of a priori reflection. Thus, any such theory transforms conceptual truths (about evaluation) into contingent claims (about human psychology) and is, as a consequence, unacceptable.

I want to emphasize that this is not just the old charge that naturalists have changed the subject—"Not talking about ethics, talking about psychology *instead*." Of course they have changed the subject, that's the whole point of their enterprise. The question is, In changing the subject, how much of

<sup>13</sup> C. L. Stevenson, *Facts and Values* (New Haven, 1963), p. 57. See also, Hume's *Treatise*, Bk. II, Pt. III, Sect. II, "of the influencing motives of the Will."

the original survives the change? A weak requirement of adequacy would be that the transformation preserve truth: what was true in the original remains true in the transformed, analyzed version. A more interesting, stronger requirement—the one I am charging gets infringed—is that in addition to preserving truth, the transformation preserve logical status: what was contingent before remain so after, what was conceptually true before remain so after. Now even if it should be true, as a matter of experimentally established psychological fact, that uniformly and without exception the attitude toward the consequence always modifies the attitude toward the cause, and never the reverse, that would not be the conceptual truth that our first evaluative asymmetry axiom is. Hence, any such psychological theory will violate the requirement that analysis preserve logical status.

Things are even worse than this. All such theories would appear to violate the more elementary requirement that analysis preserve truth. That is, the psychological analogues of our asymmetry axioms appear to be false: the facts don't bear them out. The perception (discovery, belief) that *A*, which one favors, causes *B*, of which one disapproves, is a paradigm case of the phenomena investigated recently by psychologists under the various headings of cognitive dissonance, incongruence, imbalance, etc.<sup>14</sup> Yet nowhere, to my

knowledge, does the considerable body of literature in this area, including both experimental data and theorizing, even suggest, much less support the existence of the sort of psychological asymmetries currently at issue. In short, the very area of empirical psychology, cognitive consistency research, which should have the psychological analogues of our evaluative asymmetries (were they to exist) as its private preserve, appears not to have discovered them. No doubt this could just be a matter of not having found what it has not occurred to one to look for. If so, this brings us back to the original point.

The plain fact is, philosophers don't need to wait on the outcome of any psychological experimentation to decide whether it's true that the value of consequences is always relevant to the value of what causes them, but never the reverse. That question is to be settled on the basis of a priori reflection concerning, among other things, the analysis of various sorts of non-intrinsic value, the difference between "balancing off" and modifying value, and comparable conceptual matters. Psychologists, on the other hand, must go to the laboratory to find out whether attitudes held toward causes ever modify attitudes held toward their consequences. And whatever they find out will be a contingent fact of human psychology. Hence, psychological naturalism is refuted.

*Oberlin College*

*Received November 15, 1972*

<sup>14</sup> See Robert P. Abelson, et. al., *Theories of Cognitive Consistency: A Sourcebook* (Chicago, 1968).

## V. MAKING SENSE OF "NECESSARY EXISTENCE"

BARRY MILLER

IN recent years what is meant by "necessary existence" or "necessary being" has been given a novel twist. Instead of the necessity being directly ascribed to the existence of a being or to the truth of an existential proposition, it has been ascribed to another property of such a proposition, viz., its capacity to block a certain kind of question. It is legitimate to speak of necessary existence because, if the proposition "God exists" were true, it would *necessarily* render senseless any further question of the form "Why is it that such-and-such exists?"<sup>1</sup> Thus one defect with this as a rationale for the notion "necessary existence" is that necessity can be ascribed to existence only very obliquely. Properly it should be ascribed to the effect which a certain proposition has on the posing of a certain question.

In this article I shall sketch how such a position arose, and then criticize it for selecting a proposition which cannot do even the limited job allotted to it. Subsequently I shall argue that there is a different proposition which could do such a job, and that this proposition would have the additional merit of showing how necessity could be ascribed to existence quite properly, and not merely in some oblique fashion.

### I

A familiar criticism of the notion of "necessary existence" is that only propositions can be necessary. And when the user of "necessary existence" agrees to abide by this stipulation by saying that "God exists" is a necessary proposition, he has sometimes been told that this cannot be. "God exists" is an existential proposition, no existential propositions can be logically necessary, and so "God exists" can not be logically necessary.

At this stage the theist has often pointed out that these remarks miss the point. He merely claimed

that "God exists" was necessary, not that it was logically necessary. The necessity he would propose is a non-logical one, which he calls variously "ontological necessity" or "factual necessity."

However, it is one thing to deny that all necessity need be logical, but quite another to give any substance to the notion of "non-logical necessity." This is particularly so in the present instance, for presumably "God exists" would have to be a synthetic as well as a necessary proposition. Some deny such a possibility; and Plantinga, for example, suggests two reasons for the denial. The first is that synthetic necessary propositions are such that "their denials, while logically quite consistent, are nonetheless inconceivable. And the best evidence that 'God exists' does not enjoy this characteristic is just that reasonable and intelligent people do in fact conceive its denial." The second reason is that "all other propositions said to enjoy this status describe or report relationships between possible instances of two or more properties; none of them assert that any property *has* instances or that some description actually applies to anything. 'Whatever is colored is extended' has been held to be both necessary and synthetic; that claim is not made for 'There are colored objects.' Hence, the claim that 'God exists' is a synthetic necessary proposition is implausible."<sup>2</sup>

Neither of these reasons seems to be conclusive, the first because the intuitively inconceivable is very uncertain ground to build on, the second because it is relevant only to the there-is sense of "exists" but not to its actuality sense. However, if they were accepted as conclusive, then "God exists" would not be *necessarily* true, since it would be neither an analytic proposition nor a synthetic necessary one.

It is against such a background that Plantinga and others have sought to explain the necessity of

<sup>1</sup> R. L. Franklin, "Necessary Being," *The Australasian Journal of Philosophy*, vol. 35 (1957), pp. 97-110. R. L. Franklin, "Some Sorts of Necessity," *Sophia*, vol. 3 (1962), pp. 15-24. John Hick, "God as Necessary Being," *The Journal of Philosophy*, vol. 57 (1960), pp. 725-734. Alvin Plantinga (ed.), "Necessary Being" in his *Faith and Philosophy* (Grand Rapids, 1964), pp. 97-108.

<sup>2</sup> "Necessary Being," *ibid.*, p. 104.

"God exists" in terms other than that of necessary truth. In its stead they suggest that to say "God exists" is necessary is to say that, if it were true, a certain kind of question necessarily could not be asked. Both he and Hick, for example, interpret "God exists" as being necessary in the sense that it is the point at which questions of the kind "Why is it that such-and-such exists?" cannot sensibly be asked.

This suggestion has been criticized, however, as treating the existence of God as simply a brute fact.<sup>3</sup> The theist invokes the existence of God because he is unwilling to accept the existence of the universe as a brute fact. But if the existence of God is itself a brute fact, then he is no better off for having invoked it. It would be more in accord with the principle of simplicity just to accept the existence of the universe as a brute fact, and forget about God.

Plantinga and Hick can each produce a rejoinder to this kind of objection, but neither to my mind is satisfactory. Plantinga would say that to ask why God exists "is to presuppose that God does exist; but it is a necessary truth that if He does, He has no cause; hence there is no answer to a question asking for His causal conditions. The question . . . is therefore an absurdity. A person who seriously asks it betrays misapprehension of the concept of God."<sup>4</sup> In his view, therefore, although it may be absurd to ask why God exists, it is not absurd to ask why that question is absurd. And if we can say *why* it makes no sense to ask why God exists, then it would presumably be wrong to regard God's existence as a brute fact. But I do not think that follows.

Plantinga explains that God's "beginning to exist is causally impossible, for since it is analytic that God is not dependent upon anything, He has no cause; and hence His coming into existence would be an event which would have no causally sufficient conditions. So if God does exist, He cannot . . . have begun to exist."<sup>5</sup> To reach such a conclusion, however, he needs to draw upon an unstated premiss. If God does exist, the lack of causally sufficient conditions would be insufficient to show that He did not begin to exist, unless it were

additionally assumed that every event requires causally sufficient conditions. And since that is a synthetic proposition, it would either have to be accepted as a brute fact or proved in some way. So, by invoking God, Plantinga may not succeed in avoiding the acceptance of a brute fact; and he still has the task of showing that his brute fact is preferable to the atheist's.

Hick, however, does try to do just that. He is prepared to admit that the existence of God would indeed be a brute fact, but sees it as a logically superior one to that of the existence of the universe. The latter can "function as a *de facto* terminus of explanation; but God can, in addition, function as a *de iure* or logical terminus." He explains that "the existence of God (defined as eternal, independent, and as the creator of everything that exists other than himself) . . . is an uniquely ultimate fact, behind or beyond which it is logically impossible to go; whereas the existence of the physical universe does not have this totally ultimate character."<sup>6</sup> According to this account, what marks the difference between God and the physical universe as ultimate explanations is that the former is independent and eternal and creator of everything other than himself, whereas the latter is not.

Although Hick lists three characteristics which make God superior to the physical universe as an ultimate explanation, only one of them really matters. The question of eternity is not one that need distinguish God from the universe. Even if created, the universe need not have had a beginning, but could be co-eternal with (even though dependent upon) God. Nor is being a creator a *sine qua non* for an ultimate explanation. If there were grounds for regarding the existence of the universe as a candidate for being an ultimate explanation, they would not be vitiated simply because the universe was not "creator of everything other than itself." The *ex hypothesi* absence of anything other than itself would make creation a non-question. Of the characteristics mentioned by Hick the only one that matters, therefore, is independence. It is this, apparently, which makes the difference between being a *de facto* and a *de iure* terminus of explanation. It is this which would

<sup>3</sup> Adel Daher, "God and Factual Necessity," *Religious Studies*, vol. 6 (1970), pp. 23-40. D. R. Duff-Forbes, "Hick, Necessary Being, and the Cosmological Argument," *Canadian Journal of Philosophy*, vol. 1 (1971-2), pp. 473-484. It might be noted that in his *Arguments for the Existence of God* Hick had earlier shown his awareness of this point. Cf. p. 48.

<sup>4</sup> *God and Other Minds* (Ithaca, 1967), p. 182.

<sup>5</sup> "Necessary Being," *op cit.*, p. 107.

<sup>6</sup> "Comment" (in reply to D. R. Duff-Forbes, "Hick, Necessary Being, and the Cosmological Argument"), *Canadian Journal of Philosophy*, vol. 1 (1971-2), pp. 486-487.

make it a necessary being in the special sense proposed by Hick.

However, we are entitled to ask why we should not ascribe independence to the universe rather than to some entity called God. On this point Hick's position is quite clear: "Given a beginningless universe it is logically possible to ask whether it exists in dependence upon the will of a God who is the creator of everything that exists other than himself. But given such a God, it is not logically possible to ask whether his existence is dependent upon the existence of the physical universe, or of anything else; for his relationship to anything that exists, other than himself, has been defined asymmetrically as that of creator to creature."<sup>7</sup> As is clear, the possibility of asking whether God exists dependently on something has been eliminated by definition, i.e., by stipulation. But if stipulation could be employed at that point, why could it not have been employed at an earlier point, namely, in connection with the physical universe? There is, of course, no reason. Only if the universe were demonstrably a dependent entity would we be blocked from stipulating that it is the ultimate terminus of explanation. But, while this possibility is open to some theists, it is not open to Hick, for he does not think that we can *prove* that the universe is a dependent entity.

On the other hand, we would be blocked from invoking God as an ultimate explanation only if it were perfectly clear that the physical universe required no explanation. Only because this is *not* clear can a theistic explanation begin to get off the ground. It is, of course, perfectly legitimate to answer the question "Why does the physical universe exist?" with the counter question "Why shouldn't it?" That may put the onus of proof on the theist, but it does not block entirely the possibility of a theistic explanation. While it may not be clear why the existence of the physical universe should require any explanation at all, it is equally unclear why it should not.

The reason for this is that it is not clear why the universe should not just cease to exist. And until there is clarity on that point, we are stuck with the impasse of the theist asking "Why shouldn't it?" and the atheist or agnostic asking "Why should it?" Now since this paper is partly concerned with giving some sense to the notion of "necessary existence" as a question stopper, the present impasse suggests that we might do better to ask first why the demise of the physical universe

is even conceivable, and then asking what would be required to make it *inconceivable*. If that can be answered, then the question we are trying to block could not even arise. And if we can see *why* it could not arise, we may be closer to giving a sense to "necessary existence" that, unlike those under discussion, would be more than merely stipulative.

## II

I suspect that the reason why the demise of the physical universe is conceivable is that it does not seem implausible to speak of the universe losing its existence (going out of existence). We speak of a leaf losing its color, a suit losing its shape, a man losing his speech, and so on. Even though the loss of existence is something rather more radical than any of these, there is no obvious reason why it should be ruled out. In fact, there seems good reason for saying quite the contrary, for since "exists" does not have to be predicated of any universe, the universe could quite well cease to be.

The point I am making about "exists" is simply an application of the quite general one that no predicate is necessarily predicated of anything. One may be inclined to think that such remarks are surely inapplicable where the predication results in an analytic proposition. One is no doubt thinking of propositions like "Man is an animal." But here "animal" is not being affirmed or denied of anything. What is being affirmed is that if anything is a man it is an animal. But this does not fully meet the objection, for there are some analytic propositions in which first level predication (and not merely a first level predicate) does occur, e.g., "This man is an animal." And in such cases one might well be tempted to think that "animal" does have to be predicated, and hence cannot be a contingent predicate.

But in fact there is nothing about "— is an animal" that requires its predication of any man at all. And the same is true in the case of "Fido is a dog." There is nothing about "— is a dog" which demands its predication of any dog at all. Of course it is necessary that "this man" can have "— is an animal" attached to it as a predicate; and it is necessary that "Fido" can have "— is a dog" attached to it as a predicate. But it is not necessary that "— is an animal" have any man to be predicated of, nor that "— is a dog" have any particular dog for it to be predicated of.

This may seem like laboring the obvious, but I

<sup>7</sup> "Comment," *op. cit.*, p. 486.



think it had to be said, lest the predicates in analytic propositions be regarded as necessary ones. There is indeed a sense in which they can be said to be necessary, but it is not the sense in which we are interested. Whether in analytic or in synthetic propositions, no predicate is necessary in the sense that there has to be something for it to be predicated of. And that is true no less of "exists" than of any other predicate.

No doubt the last sentence might be queried on the grounds of "exists" not being a first-level predicate. I am not moved by this objection, for I think it false. We have to distinguish two senses of "exists," one as said of individuals, the other as said of kinds, these being called by Frege and Geach the actuality and there-is senses respectively. "Exists" in the there-is sense is not a first-level predicate, but "exists" in the actuality sense is. I will not try to argue for this position, not because I think it requires no argument, but because it has already been done elsewhere.<sup>8</sup>

If it is accepted that "exists" is a first-level predicate, the question of its necessity seems to be no different from that of any other first-level predicate. And if that is so, it could make sense to say that *B* exists contingently, but it would make no sense whatever to say that *B* exists necessarily. And that last clause, surprisingly perhaps, does not differ from the position of those who *deny* that "exists" is a predicate. If it is not a predicate at all, it obviously cannot be one that must be predicated of something. And we have just seen that, even if it is a predicate, it cannot be such that it *must* have something to be predicated of. So maybe "necessary existence" is nonsensical after all, even if not for the reasons most commonly given.

Further support for this view comes from contrasting "*B* exists contingently" with "*B* exists necessarily." From the former we can infer not only that it is a contingent matter whether *B* exists, but that *B*'s existence is a contingent one. But from the latter we can infer only that it is a necessary matter that *B* exists, but not that *B*'s existence is a necessary one. Considered in itself, *B*'s existence

might well be only a contingent one, for the fact that *B* can not lose it might be explained by some determination *ab extrinseco*. If all that prevented it from being lost were merely some extrinsic intervention, then *B*'s existence would still be contingent, even though one could rightly say "*B* exists necessarily."

An obvious bar to such extrinsic determination would be if "exists" were part of the notion of "*B*," and "*B* exists" were analytic. Even this, however, would not allow us to infer that *B*'s existence was an inherently necessary one, but only that *B* was a necessary entity. The difference between the existence of *B* in non-analytic and analytic examples would not be that the first was inherently contingent and the second inherently necessary. The kind of existence would be the same, but the kind of entity having it would be different in either case.<sup>9</sup> Thus, talk of inherently necessary existence is equally inappropriate when "exists" is part of an analytic proposition as when it is part of a synthetic proposition. Or, to put it in terms used earlier, even if there were in fact something of which it had to be predicated, that still would not mean that "exists" positively requires something to be predicated of.

It is now apparent that, because "exists" is a first-level predicate, it could conceivably not be predicable of the universe, i.e., there could be no universe for it to be predicated of. For that very reason the atheist is precluded from saying that the universe cannot cease to exist. But on the same grounds he is able to block the theist from saying that *any* being (e.g., God) cannot cease to exist. Hence the impasse between them, with the theist asking "Why shouldn't the universe cease to exist?" and the atheist replying "Why should it?"

That, however, is not the end of the enquiry. Since the impasse arises from the predicative use of "exists," it might be broken if there were a non-predicative use, i.e., if "exists" could be used in such a way as not to be an incomplete expression.<sup>10</sup> It is the nature and effects of such a use that I now wish to explore.<sup>11</sup>

<sup>8</sup> Peter Geach & Max Black (eds.), *Philosophical Writings of Gottlob Frege* (Oxford, 1960), p. 146. P. T. Geach, "Form and Existence," *Proceedings of the Aristotelian Society*, vol. 55 (1954-5), pp. 251-272; "What Actually Exists," *Proceedings of the Aristotelian Society*, Supplementary Volume XLII (1968), pp. 7-16; *Three Philosophers* (Oxford, 1961), pp. 88-97.

<sup>9</sup> Similarly, the kind of attribute bespoken by "This polar bear is white" does not differ from that bespoken by "This dog is white," even though the first proposition might be analytic and the second synthetic. The difference lies not in the *attribute* of being white, but in the *kind of entity* possessing that attribute, i.e., polar bear vs. dog.

<sup>10</sup> Predicates are incomplete expressions, i.e., have incomplete sense. They make sense not in isolation, but only as part of a proposition.

<sup>11</sup> I am trying to show not that such a use is in fact required, but only what would follow, especially in regard to "necessary existence," if it were.

If there is a use in which "exists" had complete sense, obviously it would have to differ from those uses in which it has incomplete sense. The difference would have to extend at least to the point of systematic ambiguity between them. Ambiguity of that kind is of course quite common in the case of terms that can function as logical subjects or logical predicates, e.g., "man" and "shark" in the following examples, the first occurrence of each being as a logical predicate and the second as a logical subject:

"Peter is a man," "A man hit Peter."

"This fish is a shark," "This fish evaded a shark."

If "exists" were to fit into this precise pattern, however, it would be unable to retain its verbal form, and it might be suggested that the complete expression correlative to the incomplete one "exists" would be the abstract noun "existence." Now, although it is debatable whether abstract nouns *can* be logical subjects, there is no need to resolve that debate here. That "existence" is not the complete expression we seek is clear if we consider, for example, "Peter has existence." As a *predicate* "— has existence" can never be necessary in the sense we require. And even if it could be a logical *subject*, existence could no more be said to exist than whiteness could be said to be white.

Now if terms were the only expressions that could have complete sense, we might seem to have reached a dead end. But propositions also have complete sense. So, since the complete and incomplete senses of "exists" cannot be a difference between terms, there is still the possibility of it being between a term and an expression having propositional function. That is to say, there is the possibility that "exists" with complete sense would be exercising propositional function.

An obvious objection is that "exists" is logically simple. How then could it exercise the propositional function of saying something about something? That kind of function would seem to be impossible without logical complexity. *Prima facie* this seems a very strong point indeed. But to evaluate it more closely we need to consider the grounds for logical

complexity in first-level propositions as we know them; and these seem to be at least twofold:

1. *Generality of the predicate.* The one predicate is attachable to many different subjects. So *such* predicates obviously could not form a simple expression with any one subject.
2. *The need to exercise both referential and predicative functions.* First-level (but not second-level) propositions not only refer to some individual or individuals, but say something about those individuals.

As regards generality of predicate, the need for complexity would not arise if what was predicated was peculiar to a determinate individual. That of course would have to be the case if referential and predicative function<sup>12</sup> were to be combined in one logically simple expression. So the crucial question is whether logical complexity is a necessary condition for the combined exercise of those functions, or whether there could not be an expression that did exercise both functions, although in such a way that neither function could be ascribed simply to a part of the expression. No doubt it would be an extremely strange expression to us, but I do not think there is any *a priori* reason for excluding its possibility. In that case it would exercise propositional function, but in such a way that there could be only *one* thing in regard to which that function could be exercised.

At this stage I think it worthwhile stressing what I mentioned earlier in passing, viz., that "exists" used as a proposition would be at least systematically ambiguous with respect to "exists" used as a predicate (in the actuality sense). That is to say, we would now have to distinguish *three* senses of "exists."<sup>13</sup> Two have already been distinguished by Frege and Geach: these are the incomplete senses, the actuality sense and there-is sense. The third would be a complete sense, which might be appropriately called the "propositional" sense.

We ought not to be surprised at such a sense. Indeed, in the context for which it is being considered, it would be surprising were it otherwise.

<sup>12</sup> I realize that in some theories predicates are said to have reference. In distinguishing referential from predicative function I do not want to be taken as denying such a position. Therefore let me state that I am using "referential function" in the restricted sense of "function of referring to *individuals*."

<sup>13</sup> Such a multiplication of senses might move people to reach for their Occamist razors. But even Occam was prepared to admit multiplication *cum necessitate*. The necessity for the two incomplete senses has already been argued for by Frege and Geach, and that for the complete (i.e., propositional) sense would depend on whether there were some independent way of knowing that the proposition "exists" had to be affirmed. But as to the general question of multiplication of senses, this is far from unusual—witness the hoary old example of "healthy," which has at least three senses that are systematically ambiguous with respect to each other.

What is said about things that differ in kind sometimes, though not always, requires a shift in sense in the use of the one term. Thus, when we say "The acid acted quickly on the metal" and "He acted quickly to save her life" we are using "acted" in regard to things that differ in kind. We are also using it in two different, though related, senses. If the difference between God and creatures is as vast as is commonly proposed, it is quite to be expected that creature language should acquire a new, though related, sense when used of God. And my suggestion for a third sense of "exists"—the propositional sense—entirely squares with this.

As regards that sense, let me remove a misunderstanding which has actually arisen. In discussion, some colleagues gained the impression that this sense was simply a coalescing of a subject use of "exists" with a (first-level) predicate use of it. It would be as if the writing of "exists" as a proposition were merely an elliptical way of writing "exists" twice—once as a subject and once as a predicate, with these two uses being systematically ambiguous. This is not at all what I mean. A proposition understood in that way would possess only a counterfeit simplicity, whereas the proposition "exists" that I am suggesting would have a genuine one.

Let me postpone further comment on this matter until we have examined the question of whether the proposition "exists" would do what the proposition "God exists" (as used by Hick and Plantinga) fails to do, i.e., whether it would qualify as a self-explanatory proposition or simply as the expression of a brute fact. In other words, if the proposition "exists" were true, would internal evidence alone show that it could never be false, and hence that no further questions of the kind "Why is it that such-and-such exists?" could sensibly be asked?

The paradigm of an expression that internal evidence shows can never be false is, of course, an analytic proposition. But this model will not help us with "exists," because the truth of an analytic proposition is determined by its internal structure, whereas "exists," being simple, has no internal structure. Simply from a study of the proposition "exists" it could never be evident whether it was true or false. That could be determined not by internal, but only by external evidence. So in that respect the proposition "exists" is unlike an analytic proposition.

In another respect, however, it is like one, for we could say of it too not merely that it is not false but that it could not be false. This would be so only if external evidence were to show it was true.<sup>14</sup> Being simple, its referential and predicative functions are indistinguishable from each other. Hence, if true, it has a referent and affirms something about that referent. But since internal negation is impossible here, what is affirmed by such a proposition can not be negated. If we know it to be true, therefore, we can know that it could not be false; and if we can know it to be false, we can know it could not be true. In this particular respect, then, the proposition "exists" would be like an analytic proposition.

It is now time to take up a matter that I earlier postponed, viz., the unpacking of the proposition "exists." Even if such a proposition could provide a self-explanatory terminus to any series of questions like "Why is it that such-and-such exists?", its value would be rather dubious unless some sense, no matter how inadequate, could be made of it in ordinary language. It will not do merely to stipulate that "exists" is a proposition which, though simple, combines the referential and predicative functions of first-level propositions. It must be possible to indicate in some way both *what* is being referred to, and what question of truth is being raised by such a proposition.

In proposing some kind of translation I would recall four features previously noted about the proposition "exists," viz.:

1. In its simplicity it nevertheless combines the referring and predicative functions of a first-level proposition.
2. Because of its simplicity what is predicated must be *peculiar to* the referent.
3. Because the propositional sense of "exists" is systematically ambiguous with the actuality sense, probably the best (but poor) ordinary language approximation to that predicate would be "exists" in the actuality sense.
4. If, on external grounds, the proposition "exists" can be known to be true, it is evident on internal grounds that it could never be false.

Based on these four points I think that the closest ordinary language rendition of the proposition "exists" would be:

<sup>14</sup> As a matter of interest, the contingency argument as I have developed it ("The Contingency Argument," *The Monist*, vol. 54 (1970), pp. 359-373, with replies to criticism in "Naming, Predicating, and Contingency," *Sophia*, vol. 13 (1973) does show that the proposition "exists" is true.

"Something, and necessarily only one thing, exists necessarily."<sup>15</sup>

At this point I can well imagine impatient mutters of, "Why on earth didn't you use this in the first place, instead of wasting time with this mind-stretching notion of the propositional sense of 'exists'?" Very simply, the reason is that the translation is irremediably inaccurate, if for no other reason than that it uses an incomplete sense of "exists."<sup>16</sup> And, as mentioned earlier, no *predicate* can be said to be necessary in the required sense of that term. So, anyone wanting to use this imperfect translation will have to practise some kind of therapy, lest he be misled by his own language.<sup>17</sup>

So much for a translation, admittedly inadequate, of the proposition "exists." Now let me draw together some of the consequences if there were justification for such a proposition:

1. It would be both like and unlike an analytic proposition. It would be unlike an analytic proposition in that the evidence for its *truth* would be external to it. It would be like an analytic proposition in that the evidence for its *necessity* would be internal to it.
2. It would provide no grounds for an Ontological Argument, since it would not be self-evidently true but only self-evidently necessary, if true.
3. If "God exists" is, as I suggested in footnote 15, implied by the proposition "exists," then "God exists" will be a synthetic proposition, as many theists claim. At the same time, if true, it will be necessarily so. Contrary to Plantinga, therefore, it will be a synthetic necessary proposition.
4. The proposition "exists" would exemplify the

kind of necessity proposed by Plantinga and Hick. If to the question "Why does anything exist?" one were to reply "Because the proposition 'exists' is true," any further questions of this kind would be effectively blocked.

5. However, as distinct from the proposals of Hick and Plantinga, the blocking would not be effected by a brute fact posing as an explanation. Given that the proposition "exists" could be *proved* to be true,<sup>18</sup> it would be self-explanatory why the answer to "Why does anything exist?" does not permit of any more such questions. To quote Penelhum against himself, "it [the proposition 'exists'] covers not only the last 'Why?' but the next one too."<sup>19</sup> There would thus be no need to fall back on Hick's distinction between a *de facto* and a *de iure* terminus of explanation.

### III

Considered together, these points undermine the atheist's claim that to offer "God exists" as an explanation of "The universe exists" is merely to substitute one brute fact for another. If, as I have indicated, "God exists" is implied by the proposition "exists," it cannot be a brute fact, for the proposition "exists" would have to be *proved* to be true, and then the evidence for its *necessity* would be internal to it.

Though this may be of some significance, it is nevertheless only a modification and development of Hick and Plantinga. As such it is not yet clear how it escapes criticism I mentioned at the beginning of this paper, viz., that the necessity in question is, strictly speaking, ascribed not to existence but to the capacity of a certain kind of existential proposition to block a certain kind of question.

<sup>15</sup> With additional argument as to what attributes such a being must possess, it might be shown that this proposition implies "God exists."

<sup>16</sup> Lest we be unduly disturbed by such inaccuracy, it is well to recall that we are commonly guilty of inaccuracy when we translate the actuality sense of "exists" by the there-is sense.

<sup>17</sup> We need not be dismayed by this, for we are no strangers to therapy of this sort. For example, in saying "The height of the building blocked my view" it would be less misleading to have said "A building of such-and-such height blocked my view." Despite that, we are not misled by language into thinking of height rather than buildings as particulars.

<sup>18</sup> To know that a conclusion is true we need know not what the conclusion *means* but only the truth of the premisses from which it is validly inferred. In the same way we could know the proposition "exists" to be true, despite being unable even in principle to provide any adequate translation of it into ordinary language. That, of course, would not mean that the proposition was *in se* not completely intelligible, but merely that it was not completely intelligible *to us*—a distinction that could be dismissed only by begging the very question of whether there might not be a being more intelligent than man.

Thus, the possibility of proving the truth of a proposition for which ordinary language is deemed irremediably inadequate is not one to be rejected a priori: any putative proof has to be judged on its merits.

<sup>19</sup> Terence Penelhum, "Necessary Being," *Mind*, vol. 69 (1960), p. 176. This shows, too, that he is wrong in thinking that "there is no way in which the existence of any being could be held to be a fact explicable by reference to that being itself." *Op. cit.*, p. 180.

However, I think it does escape that criticism; and it is surprisingly easy to show how.

We might recall a point made by Moore in his paper "Is Existence a Predicate?" As he remarked, in asking that question we know quite well that "existence" is not a predicate, but that the point at issue is whether "exists" is a predicate. Similarly, in discussing "necessary existence" we should know quite well that the question under discussion is whether "necessarily" can ever modify "exists." Or, in other words, could an existential proposition ever be necessarily true? And what I have tried to show is the conditions under which it could, viz., if there were a third sense of "exists," the propositional sense. If true, the proposition "exists" would be necessarily true. And that would be the basis for the notion of necessary existence in the strict sense, and not merely derivative from any question-blocking capacity. On the contrary, the question-blocking capacity of the proposition "exists" would be a corollary of its being necessarily true.

Of course, all this turns on whether there is in fact any justification for asserting the proposition "exists." And to supply that justification the theist would have to show:

1. That the existence of the universe not merely permits, but positively requires that the question "Why does anything exist?" be posed.
2. That the existence of the universe cannot be the ultimate answer to that question.
3. That the ultimate answer is provided by the proposition "exists."

To establish these points with any rigor would require a further article. However, it might be fitting to conclude with the briefest sketch of just

how the proposition "exists" could provide the ultimate answer to "Why does anything exist?" Such a question would be logically unavoidable if it could be shown that the existence of some entity (say, my dog Fido) was a dependent one. And the answer to "Why does Fido exist?" would be "Because some other entity *A* (on which Fido is existentially dependent) exists," i.e., "Fido exists if *A* exists." Of course, such an answer need not be the ultimate one, for, if *A* too were an existentially dependent entity, then we could say "*A* exists, if *B* exists," and so on.

Nevertheless, to conform with our experience that Fido does exist, the series "Fido exists, if *A* exists; *A* exists, if *B* exists; ..." would need a categorical existential proposition as its terminus. But, besides being categorical, the terminating proposition would require a logical form unlike that of "Fido exists": since a proposition of such form is appropriate to an existentially dependent entity, any existential proposition like it could provide merely an apparent terminus to the series. To provide a genuine terminus the existential proposition required would have to be one appropriate to an existentially independent entity, and would be given by a proposition in which "exists" functions not as an incomplete expression but as a complete one, i.e., the proposition "exists."<sup>20</sup>

Thus, the series beginning with the proposition "Fido exists, if *A* exists" must end with the proposition "exists." It is in this way that the proposition "exists" would supply the ultimate answer to "Why does Fido (or anything else to which existential propositions like '*X* exists' are logically appropriate) exist?" But concerning the entity for which the proposition "exists" itself is appropriate, the question of why *it* exists logically could not arise.

*University of New England,  
Australia*

*Received December 22, 1972*

<sup>20</sup> I have argued at length for this in the articles mentioned in footnote 14. In those same articles I have tried to show also that the existence of Fido is a dependent one, as well as to answer the usual objections against the impossibility of an infinite series of existentially dependent entities.

## VI. KNOWING THE MEANING

BARBARA M. HUMPHRIES

The useful ways in which people ordinarily talk or seem to talk about meanings boil down to two: *having* of meaning, which is significance, and *sameness* of meaning, or synonymy.

Quine<sup>1</sup>

THERE is a third kind of useful talk about meanings: *knowing* the meaning, or understanding. In my view, the phenomenon of understanding is prior to the other two; and, when correctly interpreted, its peculiarities explain why such notions as meaning, significance, and synonymy so stubbornly resist definition.

That meanings are not entities, that knowledge of meanings is not acquaintance with objects, that "*W* means *M*" does not express a relationship between *W* and a meaning—these have been popular critical claims for some time. Nevertheless, there is a residual assumption that meanings are properties of words, that knowledge of meaning is knowledge of some linguistic characteristic, that "*W* means *M*" predicates "means *M*" of *W*. This assumption underlies most attempts to define meaning (or, the effort failing, to "repudiate" meaning). However, there is a case to be made for saying that this assumption is as misguided as the first, and it is my purpose to present such a case. What is to be argued is that a plausible sense can be given to "property-attributing" in which sense "*W* means *M*" is not property-attributing. This is to be done indirectly by proposing a sense of "property-recognizing" according to which the knowledge that *W* means *M* is not property-recognizing. An alternative account of such knowledge will be sketched along with some answers to possible objections.

I shall proceed by explaining and defending the following five claims:

1. Understanding a term<sup>2</sup> is not a matter of knowing something about the term; it is a response to a perceived term which is not a matter of recognizing a property of that term.

2. Knowing the meaning of a term sometimes occurs independently of understanding the term, but this is a derivative phenomenon which also fails to be genuinely "property-recognizing" knowledge.
3. Since knowing the meaning of term *T* is thus peculiar, "*T* means *M*" is not to be understood on the model of "*Fx*" or "*Fxy*," and in that sense, meaning is not a property of terms.
4. There are aspects of understanding term *T* which are not adequately illuminated by construing understanding on the usual model of practical know-how.
5. Statements of the form "*T* means *M*" are understood and believed by taking one's response to '*M*' to be a sample of the correct response to *T*.

### CLAIM I

In its primary sense "understanding" labels a phenomenon which, like fright, excitement, and recognition, occurs at specific times in response to perceived objects. It contrasts with phenomena like intention and hope which do not ordinarily occur momentarily in reaction to perceived objects. In a secondary sense, one can be said to understand a word even when not perceiving it, just as it can be said that Smith recognizes Russian airplanes or that flowers remind Smith of death, even when Smith is not actively recognizing or being reminded. I will call the secondary dispositional phenomenon the "latent" version and the basic occurrent phenomenon the "active" version. (It should be noted that some phenomena which are not responses, as defined below, have both an active and a latent form—e.g., wishing.) I call understanding a "response" because that phenomenon cannot be attributed to a person with respect to object *O* if his perception of *O* never results in the occurrence of the phenomenon. So, for example, trusting *O*, believing something about *O*, and

<sup>1</sup> W. V. O. Quine, *From a Logical Point of View* (New York, 1961), p. 11.

<sup>2</sup> I use "term" as logicians have always used it and not as synonymous with "word."

hoping for *O* are not responses since it is perfectly possible that Smith could, e.g., trust *O* even though he frequently perceives *O* and yet that perception does not bring about his trust. However, it is not possible that Smith understands or recognizes or is frightened by *O* if Smith frequently perceives *O* and yet never reacts with understanding, recognition, or fright; so, these phenomena (i.e., the phenomena indicated by such words when the foregoing is true) are responses. More strictly, we might say:

*R* is a response to *O* iff for anyone who has *R* to *O* (latently, if the distinction is applicable) there is some aspect *A* of *O* such that there could not be anyone who (a) had *R* to *O* (latently, if the distinction is applicable) under aspect *A*, and (b) frequently perceived *O* as *A*, and (c) never on those occasions had *R* (actively, if the distinction is applicable) to *O*.

I add the qualification about aspects since we might want to say that snakes frighten Smith even if fright does not result from his perception of a snake when he sees it as a rope.

On this definition, believing Nixon to be a Californian is not a response, since it is not the case that if I believe this of Nixon, then there is some description of Nixon such that: anyone who believed Nixon (under that description) to be a Californian must regularly react to Nixon, so recognized, by believing that he is a Californian. But being frightened by snakes is a response since no one could be frightened by snakes qua *A* and frequently perceive snakes qua *A* and never react with fright. And of course, understanding is a response since if you understand *T*, there is an aspect of *T* such that no one could frequently perceive *T* qua *A*, always fail to react with understanding, and yet be said to understand *T*.

Now among responses, there is an obvious intuitive difference between responding to a perceived object by recognizing that it is round or that it belongs to Tom, and responding by coughing, or being pleased, or thinking of Stalin. Intuitively, the latter responses do not seem to be instances of knowing or recognizing something about the perceived object. I shall call reactions of the first type "property-recognizing" responses ("PR" responses).

It seems fairly evident to me that understanding a term is a response of the second, non-PR kind, in spite of the fact that it may be called "recognizing" or "knowing" or "realizing" the meaning of the

term. Indeed, I think that understanding a term is rather like<sup>3</sup> thinking of *X*—what used to be called "having the idea of *X* come to mind." However, I am aware that parts of this view are somewhat out of favor, and so I shall now argue for the claim that understanding a term is a non-PR response; to do this, I must first give a criterion for this intuitive notion.

I want to rebut the person who says that when he understands term *T*, he is recognizing a property of *T*—namely, that *T* means *M*. So let us consider the general case of a person, *S*, who, responding to object *O*, says that he recognizes that *O* is *F*. I offer the following criterion as a necessary and sufficient condition for response *R* to be non-PR even though *S* attempts to characterize *R* as "taking *O* to be *F*." (I here intend "taking" to signify the version of recognition that does not imply that *O* really is *F*.)

- (a) The alleged content ("*O* is *F*") can be relativized to *S* (the responder) if it is not already so relativized, and
- (b) There is some aspect *A* of object *O* such that *O* would be *F* to anyone who frequently perceived *O* as *A* and always on those occasions responded with *R* (allegedly, taking-*O*-to-be-*F*).

The intuitive idea here is that in non-PR responses any alleged property of object *O* is really constituted by the response itself. A simpler version of (b) would be:

- (b') *O* would be *F* to anyone who frequently perceived *O* and always responded with *R*.

I add the qualification about aspects in (b) with the following kind of case in mind. We might want to say that being frightened by Fido is non-PR, but that Fido would not be frightening to Smith even if Smith was regularly frightened at night by a dark shape in the garage (which turns out to be Fido).

Taking *O* to be poisonous (to someone) is genuinely PR since (no matter whether *O* is perceived as a green pill, a round object, etc.) *O* could fail to be poisonous to someone even if he regularly took *O*, as so perceived, to be poisonous. On the other hand, insofar as pleasure is a response, "taking something to be pleasing" has two senses, one PR and one non-PR and the criterion shows this. Suppose someone thought that the response that we ordinarily call "being pleased by *O*" is really a PR response—"taking *O* to be pleasing." Let us suppose that *O* is a rose. We see that for some aspects of that rose (e.g., its being the red flower in that

<sup>3</sup> See Claim 5.

vase), it is true that the rose would be pleasing to anyone who frequently perceived it as the red flower in the vase and who always had that response when he did so perceive the rose. In other words, that response is such that whoever regularly has it in response to his perception of *O* as *A* would be someone to whom *O* is *F*. So, that response is non-*PR*. However, suppose Smith, to whom the rose is pleasing in the way just described, comes into the room blind-folded and recognizes the rose by touch; the touching of the rose does not give him the response we have just been discussing, but he thereby recognizes the rose as something which does (latently) please him. Here is the sense of "taking *O* to be pleasing (to *S*)" that indicates a genuinely *PR* response, a case of recognizing something about *O*, namely, that *O* is an object which is pleasing (to *S*, to me, etc.). For the response that Smith had when blind-folded is such that someone might regularly have that response to (some aspect of) *O* even though *O* is not pleasing to him. And so, this sense of "taking *O* to be pleasing" indicates a genuinely *PR* response.

How does understanding fare by this criterion? It is non-*PR*. Suppose someone thought that understanding *T* is a matter of (*PR*) taking *T* to mean *M*. It cannot be. For there is some aspect *A* of *T* (its spelling or phonemic constitution, say) such that *T* would mean *M* to anyone who frequently perceived *T* as *A* and always on those occasions had that response, i.e., if you regularly "take" *T* to mean *M* then *T* means *M* to you.

Let us consider some possible difficulties in the criterion that I have proposed. As might be expected, there are a number of marginal cases which, for one reason or another, are not as clear as those in which I am interested. I shall say a few words about them, but my general advice to the reader who finds them troubling is to think of the criterion as amended to exclude such cases. For, understanding does not fall into any of the categories to be discussed and so nothing would be lost by distinguishing *PR* from non-*PR* only among cases which do not fall into any of the categories below.

(a) Moral judgments. It may be doubted whether or not *O* would be good, to Smith, if Smith regularly failed to take it to be so. This kind of case is, however, a merely apparent difficulty. Moral judgments are not typically responses—Smith's evaluation of *O* does not depend on how he reacts to *O* when he perceives it. (b) Purely internal states. Would a pain be throbbing to anyone who regu-

larly took it to be so? Again I shall excuse myself from debating this problem on the ground that this does not involve a response to a perceived external object. (c) Aesthetic judgments and some sensory judgments. These are responses, so we may ask "Is there some response *R* (which might plausibly be called 'taking the statue to be beautiful') such that the statue would be beautiful to anyone who regularly responded with *R* when he perceived (some aspect of) the statue?" Or: "Is there some response *R* (which might plausibly be called 'taking the wine to be sweet') such that the wine would be sweet to anyone who regularly responded to the wine with *R*?" It seems to me that it is difficult to answer and that those who do answer will probably disagree. This shows that there is some unclarity in our ideas about these matters. The criterion will not resolve this, but it will correctly separate those who think of such responses as the recognition of something objective about the object from those who do not. (d) Appearance predicates—e.g., looking red or tasting sweet (as opposed to being red or sweet). Is there a response properly called "taking *O* to look red" such that *O* would look red to anyone who regularly "took" it to look red? Two remarks: (i) This is in part the problem of whether there is a looks-red reaction, a sort of presentation, which is independent of any taking or judging of an object to be really red or not red. (ii) This problem is further complicated by the reference to perceived aspects in the criterion; one has to decide what to do about the proposed substitution "*O* would look red to anyone who frequently perceived *O* as red and . . .".

## CLAIM 2

We have seen that "taking" *T* to mean *M* is no more *PR* than is being pleased by the rose. Indeed, it is less so. For there is no version of taking-*T*-to-mean-*M* (to me) which is *PR*. In the case of the rose, there was the possibility of recognizing by touch that this is something which is (latently) pleasing (to me); for it was not the case that anyone who regularly had such a response to a certain aspect of the rose is really pleased by the rose. However, let us consider an analogous situation with respect to taking-*T*-to-mean-*M*. Suppose someone is to give a speech; I am watching from a distance, and my companion, who is very well-informed about the speaker, says "I don't know exactly what the first word out of his mouth will be, but I know it will be a word which means money." The speaker begins; I hear the sound,



but cannot make out the words; nevertheless, I take the first word, whatever it is, to mean money. I have not understood the term, for (as is not the case with pleasure) a person is usually said to understand *T* only when he responds to a certain aspect of *T*; and that aspect of *T* is not "being the first word in this speech." However, my response is such that for some aspect *A* of *T*, *T* would mean *M* to anyone who did frequently perceive *T* as *A* and who always on those occasions had that response. So my response here is also non-*PR*. In the case of pleasure there is such a thing as (*PR*) recognizing that *O* is (non-*PR*) pleasing to me without being pleased. But in the case of meaning, the response I have when I "realize" that *T* means *M* (to me) without understanding is also non-*PR*. And this is as it should be. Intuitively, the only difference in response between me and a person standing close enough to the speaker to understand the word is that our responses were triggered by different aspects of the word; if one is *PR*, then the other should be too. On the other hand there is a great intuitive difference in response between the man who is pleased by the rose and the man who simply recognizes it as something which is pleasing to him. Therefore, taking *T* to mean *M* (to me) is non-*PR* even when it is not understanding.

#### INTERPRETIVE DIGRESSION

It will be seen that one effect of the discussion so far has been to place the hearer's (or reader's) reaction at the core of the notion of meaning. Instead of taking meaning to be determined by speakers' practices or beliefs or intentions and asking hearers to understand by discovering these, I wish to take "means *M* to *S*" as central and let the speakers do their meaning by intending that *T* will mean *M* to those who correctly understand. I have argued for this shift indirectly in that I have focused on hearers and argued that the understanding response is not *PR*. If it is not *PR* then *a fortiori* it is not a matter of recognizing *T* to be a term which is treated thus-and-so by speakers. I shall not here rehearse the notorious difficulties that are encountered in any attempt to specify exactly what property of the word an understander is supposed to recognize. These difficulties are all supporting evidence for my case; for, what I have tried to do is to give a sense and a defense for the view that understanding is not *PR*, and thereby account for those failures of definition.

#### CLAIM 3

Hypothesis: If a sentence of apparent form "*Fx*" expresses genuine knowledge or belief about *x*, then it is possible to respond to *x* by taking *x* to be *F*, in a genuinely *PR* sense. We have seen that every apparent recognition that *T* means *M* is merely apparent and not genuinely *PR*. In conjunction with the hypothesis just given, it follows that "*T* means *M*" does not express a genuine belief about *T*. In a parallel sense, meaning *M* is not a property of *T*; whenever you think you are recognizing that property you are in fact having a non-*PR* response. Believing that *T* means *M* is thus like being pleased by roses and not believing that *X* is red; it cannot be knowledge in the usual sense; it is, rather, a kind of latent, dispositional state the active manifestation of which is a non-*PR* response.

However, to support such a view, much more is required than what I have done so far. Since I have claimed that "*T* means *M*" does not predicate a property of *T*, I must (a) consider the view that understanding is a kind of know-how; (b) explain what is involved in understanding, accepting, and rejecting sentences of the form "*T* means *M*"; (c) give an account of the operation of truth-functional connectives applied to such sentences; and (d) give an account of what it is to understand, accept, etc., intentional sentences involving "*T* means *M*" (e.g., "Smith believes that *T* means *M*"). In the remainder of this essay I shall discuss the first three matters; I shall not discuss the fourth since I believe that intentional operators like "believes that" work in something like the special way "means" works, and this complicates the issue.

#### CLAIM 4

Given that understanding is not *PR*, it is tempting to regard it as a brand of know-how. For knowing how to use a knife or how to dance the eightsome reel also tend to resist classification as knowledge that. . . . And no doubt there is an intimate connection between understanding and some kind of know-how. Nevertheless, I am inclined to think that we cannot identify the two, chiefly because know-how is not a response and understanding is. Know-how is not a response because it does not have an active form such that one cannot have latent knowledge of know-how if one frequently comes across the relevant object and never actively responds (see Claim 1). For example, consider knowing how to ride a bicycle. What would be

the active form of this know-how which must occur if one is to have the latent form? Not riding-the-bicycle; for it is perfectly possible to know how to ride, frequently perceive *O* as a bicycle, and never ride. Not recognizing-*O*-as-an-object-for-which-I-have-some-know-how; for (a) the know-how might be merely about fixing the bicycle, not riding it, and (b) it is perfectly possible to know how to ride, frequently perceive bicycles, and never to have the thought cross one's mind that one has know-how about it. And not recognizing-*O*-as-an-object-for-which-I-have-know-how-*K*; for (a) need not be able to formulate one's know-how in a *PR* way, and (b) even if this is somehow a non-*PR* recognition, it is possible to know how to ride, frequently perceive bicycles, and never have it (non-*PR*) occur to one that one knows how to ride, etc.

Understanding, however, is a response, requiring an active form. Suppose understanding *T* were a matter of knowing how to act in various circumstances when someone says *T*. What is the active form? Not acting in the correct way since one needn't do that in order to actively understand or to have that know-how. And not recognizing-*T*-to-be-something-to-which-I-know-how-to-react, for reasons parallel to those given about the bicycle; latent know-how does not require any such response; and neither understanding nor know-how requires such responses if they are *PR*. Suppose instead that understanding *T* were a matter of knowing how to have the active non-*PR* response that constitutes active understanding; the nature of understanding demands that this response should actually occur, the nature of know-how does not.

Thus there is difficulty in supposing understanding to be know-how because know-how is not a response. Nevertheless, it may be possible to identify active understanding with something closely related to know-how—something I call “realizing how to.” This is the kind of response that occurs when, for example, having pondered a complex tool for some time I suddenly realize how to use it. This may not be expressible as knowledge—that, and it may or may not be an “application” of previously existing know-how. This notion needs clarification, but it seems conceivable that active understanding could be a matter of realizing how to act, given that *T* has been uttered.

## CLAIM 5

### A. The Correctness Phenomenon

So far, I have emphasized the similarities between understanding and other non-*PR* responses. However, there is a glaring exception to this parallel. It is this. There is a description of the understanding response and a sense of “correct” such that anyone who recognizes taking-*T*-to-mean-*M* to be correct must take *T* to mean *M*; if Smith sees “chai” and recognizes that taking-“chai”-to-mean-tea is correct, then Smith must be taking “chai” to mean tea; if he knows the correct way to understand “chai,” he understands it that way himself. This is entirely atypical of non-*PR* responses.<sup>4</sup> There is no sense of “correct” such that if you take it to be correct to respond to *O* with fright or disgust, you are automatically frightened or disgusted. However, there is no such thing as thinking “There’s a word which I know to be correctly understood by taking it to mean tea; but will I or won’t I understand it that way?” Yet it is easy to imagine such thoughts for most non-*PR* responses. The responses that typically show such a feature are the *PR* ones, where the sense of “correct” is close to that of “true.” For example, if I recognize that it is correct to take (*PR*) *O* to be poisonous, then I do take *O* to be poisonous.

This phenomenon is related to formulae like “If I believe that ‘*p*’ is true, then I believe that *p*” and “If I believe that knowing the color of *O* is knowing that *O* is red, then I believe that *O* is red.”

This phenomenon is mysterious enough in the case of *PR* responses; for, either knowing that *R* is correct is, in some sense, not really knowing something about *R*, or else it is a queer kind of knowledge which logically requires one to have *R* if a certain thing is known about *R*. Mysterious or not, it may now be very tempting to suppose that the “correctness-phenomenon” shows that understanding is really a typical case of *PR* recognition. However, I am sufficiently impressed by the distinction drawn by my criterion above, to resist any assimilation of understanding to ordinary knowledge. I shall instead account for the correctness phenomenon in a different way.

<sup>4</sup> Atypical, but perhaps not entirely unknown. Consider being reminded of Stalin by a hammer-and-sickle design. This is a non-*PR* response. However, would it be possible for someone to recognize the design as something which is supposed to be a reminder of Stalin and yet not be reminded of Stalin? The natural answer is “No” unless one has been corrupted by Fregean semantics. This suggests that the correctness phenomenon is not peculiar to *PR* responses but instead depends on whether the response (*PR* or not) is intentional (*i.e.*, of or about something).

Conjecture C: There is a state  $R$  which is (actively) had by anyone who takes (non- $PR$  active)  $T$  to mean  $M$  (in any context), or who takes (actively) the taking-of- $X$ -to-mean- $M$  to be  $F$  (for any  $F$ ).

In other words,  $R$  is a state which one has when one understands ' $M$ ', when one takes the first word on the page to mean  $M$ , when one takes it to be correct (or fun or unusual) to take  $T$  to mean  $M$ , and so on. (I think of  $R$  as similar to having an idea of  $M$ , i.e., as an intentional state which is "of  $M$ " but whose mode is not yet specified; however, this interpretation is not necessary to this point; one might, e.g., think of  $R$  as an (active) realization-how-to-act-if-circumstances-are- $C$ ).

In all those kinds of cases  $R$  occurs, though there will of course be some differences among the cases. The result is that whenever one responds to  $T$  by taking taking- $T$ -to-mean- $M$  to be  $F$ , one is *ipso facto* (in part) responding to  $T$  with  $R$ , and thus taking  $T$  to mean  $M$ .

It may appear that this will account nicely for the correctness phenomenon while making a shambles of all other "takings" about  $R$ . In the extreme case, it would seem to follow that if one takes taking- $T$ -to-mean- $M$  to be *incorrect*, one is *ipso facto* responding with  $R$ , hence taking  $T$  to mean  $M$ . And how could it be that if one takes it to be wrong to understand  $T$  by way of  $R$ , one not only can but must be understanding  $T$  that way?

To explain this let us consider some further features of understanding. Understanding is a "mis" phenomenon like calculating, handling, applying, etc.; it is possible to mis-do the thing, so that in one sense one does<sup>1</sup> it, while in another sense one has not done<sup>2</sup> it—because doing<sup>1</sup> it is not the correct way. Example: A certain doctor treats (latently, therefore actively) burns. He does it by massaging the burned area. So in a sense he does not really treat<sup>2</sup> burns—he does something quite wrong which could best be called "mis-treating" these cases. In other words, he treats<sup>1</sup> burns, but he does not treat<sup>2</sup> them. So too, the student who takes "validity" to mean truth understands<sup>1</sup> the word that way, and it does mean truth to him; but he is misunderstanding and therefore not understanding<sup>2</sup>.

Let us return to the correctness phenomenon. The man who takes taking- $T$ -to-mean- $M$  to be an incorrect response is understanding<sup>1</sup>  $T$  with response  $R$ , and if he does so most of the time he may be said to (latently) understand<sup>1</sup>  $T$  that way. But he also realizes that this is a misunderstanding

and he may very well understand<sup>1</sup>  $T$  in some other way as well; indeed, he may understand<sup>2</sup>  $T$  (i.e., understand<sup>1</sup> correctly). There is actually nothing strange about saying "I'm sure I am misunderstanding that word"—i.e., nothing strange in understanding<sup>1</sup> while thinking that one is not understanding.<sup>2</sup> Misunderstanding is not only compatible with, but requires understanding<sup>1</sup>; and the same is true of thinking-one-is-misunderstanding—i.e., taking taking- $T$ -to-mean- $M$  to be incorrect. Similarly, the surprising statement "If he takes taking- $T$ -to-mean- $M$  to be correct, then he takes  $T$  to mean  $M$ " can be defused by (a) reading it as "If he takes  $R$  to be understanding<sup>2</sup>, then he understands<sup>1</sup>," and (b) adopting conjecture  $C$  according to which understanding<sup>1</sup> is automatically included in taking  $R$  to be understanding<sup>2</sup>.

This discussion needs to be supplemented by a clarification of "correct"; and it also needs to be supplemented by a clarification of the notion of taking- $O$ -to-be- $F$  by, in part, having  $O$ . Some of this will be done below. My purpose here has been to circumvent an objection to the non- $PR$ -ness of understanding—an objection based on the correctness-phenomenon and dedicated to the assimilation of understanding to  $PR$  knowledge. I have argued that the correctness-phenomenon can be explained without reverting to "It's a miracle, about which we know only that it occurs typically in  $PR$  responses." I have given an alternative explanation by way of conjecture  $C$ .

#### *B. How to understand, accept, and/or reject "T means M."*

If understanding is, as I have argued, a non- $PR$  response, we are left with the problem of explaining the unquestionable fact that sentences of the form " $T$  means  $M$ " are understood and accepted, and of doing so without supposing that they attribute some property to  $T$ . According to conjecture  $C$ , (actively) understanding " $T$  means  $M$ " involves having  $R$ , which is the non- $PR$  response that constitutes understanding ' $M$ ' or  $T$  (when it occurs in reaction to certain aspects of the term). But what role does this response play in understanding " $T$  means  $M$ "? How is it different from its role in understanding " $A$  is  $M$ "?

Consider the case of Smith, who is told that a certain term (which is either presented in quotation marks or specified by description) means dogs—i.e., someone says to him " $T$  means dogs." If Smith understands, he now has a non- $PR$  response to "dogs," and this is the very response he is to have

when he perceives term  $T$ . This is what Smith wants if he asks someone about the term—he wants to undergo just what he should undergo if he saw  $T$ . If you like, he wants to know the response, but only in the sense in which someone might want to know happiness or want to be acquainted with grief. That response is communicated in something like the way diseases are communicated—it is given to Smith. In general, when “ $M$ ” is so used (i.e., in “means  $M$ ” or in “taking  $X$  to mean  $M$ ”) it functions to give the understanding response. It is as if “ $T$  means dog” were saying, “You are about to have the response appropriate to  $T$  (now keep on understanding as usual): dog.”

Why does “means  $M$ ” function this way? Why is it that, when we ask after meaning, what we want is a sample response? I don’t know, but three things seem clear enough. First, this kind of knowing-by-having is so fundamental that many people consider it to be the primary mode of knowing (cf., “I want to really know (what it’s like to) love.”) Secondly, the usefulness of such “knowing” by having need not depend on there being something about  $R$  which can be recognized only if one has  $R$ . For, third, it is possible that one wants such “knowing” for its effects; just as no amount of thinking about a disease will give one immunity, perhaps what we are after when we ask about meaning can only be brought about if the response is sampled.

The upshot is that in understanding “ $T$  means dogs,” the response to “dogs” must be taken as a sample of the response to  $T$ . However, this “taking” must be of a somewhat peculiar kind since the statement does not describe the response but gives it to the understander. My best guess at the moment is that understanding “ $T$  means  $M$ ,” i.e., having- $R$ -and-“taking”- $R$ -as-a-sample-of-the-response-to- $T$ , is a matter of realizing how to adopt or reject the latent understanding of  $T$  advocated by “ $T$  means  $M$ .” (See Claim 4 and definition No. 3 below.)

Returning to the attempt to give an account of the basic operations on “ $T$  means  $M$ ,” we may say:

- (1)  $S$  (actively) accepts “ $T$  means  $M$ ” iff if  $S$  were to perceive  $T$  as  $T$  and he has not forgotten or changed his mind,  $S$  would respond to  $T$  by (actively) taking  $T$  to mean  $M$ .
- (2)  $S$  (actively) rejects “ $T$  means  $M$ ” iff if  $S$  were to perceive  $T$  as  $T$  and had not changed his mind or forgotten, then  $S$  would

either not respond to  $T$  by taking  $T$  to mean  $M$  or he would regard such a response as misunderstanding.

- (3)  $S$  understands “ $T$  means  $M$ ” iff  $S$  realizes how to accept/reject “ $T$  means  $M$ ”. (I here allude to the unclarified notion of realizing-how-to).
- (4)  $S$  believes “ $T$  means  $M$ ”—  
if active,  $S$  accepts “ $T$  means  $M$ .”  
if latent,  $T$  means  $M$  to  $S$ —i.e.,  $S$  regularly responds to  $T$  with taking- $T$ -to-mean- $M$ .
- (5)  $S$  (latently) knows the meaning of  $T$  iff  $S$  (latently) understands  $T$  correctly.
- (6)  $S$  knows that  $T$  means  $M$  iff  $T$  means  $M$  to  $S$  and it is correct for  $T$  to mean  $M$  to  $S$ .
- (7)  $S$  accepts “ $T$  does not mean  $M$ ” iff  $S$  rejects “ $T$  means  $M$ .”
- (8)  $S$  does not accept “ $T$  means  $M$ ”: Ordinarily — $S$  rejects “ $T$  means  $M$ .”  
As logicians read this:

Even if  $S$  were to remember this and did not change his mind and were to perceive  $T$  as  $T$ , he might not respond by taking  $T$  to mean  $M$ ; and if he did have that response, he would take it to be a misunderstanding.

### C. Truth-functions on “ $T$ means $M$ .”

Truth-functions on “ $T$  means  $M$ ” are, of course, defined as usual; “ $T$  means  $M$  or  $T$  means  $N$ ” is true iff at least one of the disjuncts is true, etc. Two kinds of special problems may arise. (1) How do truth-functions work within “ $T$  means  $M$ ”—e.g., “ $T$  means  $M$  or  $N$ ”? (2) How do we explain understanding of truth-functions on “ $T$  means  $M$ ” (as opposed to explaining the truth conditions for such complexes)? I shall not discuss either question at length. With respect to (1) it should be noticed that there is ambiguity in both “ $T$  means  $M$  and  $N$ ” and “ $T$  means  $M$  or  $N$ .” For example, the first may be like “Bachelor” means male and unmarried,” or it may be like “Funny” means amusing and peculiar.” The right answer to (2) is that if you understand “ $T$  means  $M$ ” and you understand truth-functional connectives, then you understand truth-functional compounds about meaning. The reason that there may appear to be a question is that I have insisted on sample-reactions, so that it may seem that understanding “ $T$  means  $M$  or  $T$  means  $N$ ” requires one to take

*T* to mean *M*. However, this problem is avoided (postponed) by the "realizing-how" in B: (3).

#### CONCLUSION

I have defined a class of phenomena which I call "responses" and I have argued that understanding term *T* is, like fright and recognition, a response and therefore not know-how. I have further defined a sub-class of responses which I call "non-property-recognizing" and argued that understanding term *T* belongs to this sub-class. Since understanding *T* is non-*PR*, a sense can be given to the claims that "*T* means *M*" does not specify a property of *T* and

that attempts to define such a property are therefore bound to fail. I have offered an account of what it is to understand and accept "*T* means *M*" which is compatible with the foregoing; this account depends on the conjecture that the same response always occurs when '*M*' is understood although that response is, in certain contexts, treated as a sample; the same conjecture serves to explain a peculiarity which I have labeled "the correctness phenomenon." These claims have a direct bearing on certain problems about analyticity, and I believe that they are indirectly relevant to some problems about psychological statements.

*Wayne State University*

*Received December 1, 1972*

## VII. THE ANALOGICAL ARGUMENT FOR KNOWLEDGE OF OTHER MINDS RECONSIDERED

THOMAS M. OLSHEWSKY

THE analogical argument has served as the chief whipping boy for both sceptics and therapists in discussions of the problem of other minds. It is my hope here to restore it to a status of respectability. In doing so, I must argue that it is not properly an argument; nor is it appropriately concerned with knowledge or with minds. Indeed, I must contend that, as it has traditionally been dealt with, it is not even analogical. This appears to give away everything, but in the light of such criticisms I hope to offer a reconstruction of the matter that will prove more fruitful than the traditional formulations.

### I

My sense of negation in the matters regarding argument, knowledge and mind is not so much a denial of concerns as a call for refinement. We have come to speak rather loosely of the role of argument in human inquiry, often obliterating differences between discovery procedures and justification procedures. Nor do we usually draw the distinction that I propose to draw between knowledge and understanding, but I think it will prove to be a key for unlocking some pathways shut tight by subjective-objective perplexities. My point regarding minds is simply that I think when we talk about other minds, our concern is really with other persons, with knowing others as persons.

Arguments are the giving of reasons (or the reasons so given) for conclusions held. Conclusions are themselves arrived at through some process of uncovering, exposing, revelations or insight that we usually call a discovery procedure. Now, there is much debate over whether this procedure is a method; and, if so, whether it can be explicated as a method or a variety of methods, be they induction, abduction, retroduction, or seduction. But whatever these procedures are, they are not justification procedures. Justification comes his-

torically after the conclusion is arrived at, and is an attempt to set forth rational conditions for an acceptance already established. The mathematician has his theorem before he undertakes his proof; the physicist has his hypothesis before he undertakes his experimentation. Whatever the discovery procedures for arriving at the theorem or hypothesis, they come before the conclusion; whatever the justification procedures to give them logical foundations, they come after the conclusion.

There may be no logical reason why the discovery procedures and the justification procedures for a given conclusion could not be one and the same. Indeed, on a second-hand basis they frequently are. After the initial discovery has been justified by some good reasons, the best way for the discoverer to initiate others into his discovery may be just by the very argument of those good reasons. For them, the history of the discovery and the logic of justification are one and the same. This is made possible by their guide who understands why they are following this procedure and how it will lead them to the conclusion he wished to convince them of. But without the guidance of one who already understands where the goal is, they would not likely discover systematically sound reasons for arriving at it.

The role of analogy in inquiry is as a heuristic in discovery rather than as an argument in justification. In mathematics, where analogy appears as a proportion within a single definite system, it appears with at least one element unknown, and thus as a problem to solve, not as an argument in justification. Where the proportion links two systems—e.g., ratios of algebra to ratios of geometry—this may call attention to possible coordination between the systems, but does not itself argue for a linking of the two. That requires further definitions and/or axioms, if not the development of a totally new system. In a somewhat similar manner, analogies in physics argue

for nothing but offer new ways of conceptualization, the models of which are then tested by deduction and experimentation. The role of analogy in poetry is not different in kind from these but is again calling for seeing matters in a new way. That which sets poetry off from science in human inquiry is that the former calls for appreciation where the latter calls for justification. The role of analogy comes before in both, as a device for discovery.

A criticism often raised of the analogical argument for knowledge of other minds is that, unlike other arguments involving analogy, the conclusion is impossible of confirmation on any grounds other than the analogy itself. So long as analogy is itself regarded as an argument form, this criticism seems limiting, but hardly devastating. The analogy can find no corroboration from other investigation, but it can still be a sufficient argument. Once we recognize that analogy, in any human inquiry, is a heuristic for discovery, not a good reason in justification, the criticism can be seen to be disastrous. It makes clear, as the matter has been traditionally formulated, that the analogical approach offers no justification for the claim to knowledge of other minds at all.

While the negative import of this analysis is to eject analogy from any claim to a role in argument, its positive import is to call for some refinements and revisions in how we construe human inquiry. An impediment to such a project is our continuing failure to sort out ways of understanding from what is known. In the practice of inquiry, the two are not readily separable, and any adequate theory of discovery procedures must see them as interdependent and interactive. By virtue of our ways of understanding, we sort out what is relevant and what is irrelevant to our inquiry, and thus what counts as a discovery. By virtue of what we know, we construe this or that understanding as more or less adequate as a basis for our inquiry. Yet, understanding can be construed as the aspect of inquiry concerned with manners of seeing *as*, while knowledge can be construed as matters of seeing *that*. All seeing involves both seeing *as* and seeing *that*, and no human inquiry comes to rest in a conclusion without both a conceptual framework and an empirical content. (Herein lies the nub of controversies about a synthetic *a priori* and about the relation and differences between analytic and synthetic propositions.)

The apparent interchangeability of uses of "know" and "understand" in English (contrast "erkennt-

nis" and "verstehen" in German; "ἐπιστήμη" and "οὐσιολογία" in Greek) makes the sorting out all the more difficult. The distinctions I have here suggested between "ways" and "whats," between "as . . ." and "that . . .," between "conceptual" and "factual," are suggestive, not definitional. Understandings are usually stated as categorizations, characterizations, or relations ("I know him *as* a king, *as* tall and handsome, *as* the father of my spouse"); knowings are usually stated as propositions ("I know *that* the train did not arrive on time"). The former are cast as the way one thinks about things; the latter as the way things are. They are thus making different kinds of claims and subject to different kinds of justification. The one focuses on interpretation (and is better or worse, more or less appropriate, more or less adequate, than other interpretations); the other focuses on assertion (and is true or false). The visual imagery is inviting, since it is in the light of our understandings that our observations are illuminated in one way or another, and without some such light, we cannot see at all. One man may "see" a glass of water as drinkable liquid; another, as  $H_2O$ . The differences in the ways of understanding will influence what they observe about water, what they may do with it, what they may discover about it. (The difficulties of the understanding/knowing divide are exposed in the man who "sees" the water as  $HCl$ , which may be a matter of his not knowing the differences between the formula for water and the formula for salt. But this is a "fact" in the broader conceptual framework of the compounds of molecular chemistry, and his mistake could lead to a variety of intertwined misunderstanding and misinformation.)

This construal of ways of understanding as manners of seeing-as gives us a locus for analogy in inquiry. Analogies are invitations to ways of understanding, a call to see relationships and similarities not noticed before. As such they bring into focus certain features and functions that might otherwise remain at the level of subsidiary awareness. More powerful analogies carry one beyond the understanding at first attended to into further correlations not envisioned at the outset. From the base of such models for understanding, implications can be drawn and applications undertaken which serve to determine whether the understanding is a more or less plausible one.

Once we see analogy in the context of discovery rather than of justification, in the role of understanding rather than of knowledge, then my claim

about persons rather than minds becomes capable of articulation. The issues are not whether others have minds (contra solipism), nor whether we can know them (contra scepticism), nor even how we can know them. The issues are rather on what grounds do we understand persons as persons, not just as things, and what do such understandings imply for what we observe about them and how we relate to them. Clearly, one way of understanding them is in terms of a mind/body dualism, a way of understanding persons that led to posing the problem of other minds in the first place. But even without that problem, it seems to me we are less concerned with knowing other minds, or even other mental entities or events, than we are concerned with a way of understanding persons that will lay a basis for knowing what a person is doing and what is happening to him. To such an end, analogy may prove extremely helpful, but not in the form of the traditional argument.

## II

"Analogy," from the Greek "*analogia*," finds its etymological roots in the form of a proportionality, and so it was exposed and employed by Plato and Aristotle.<sup>1</sup> Our model then is of a mathematical proportion:  $A:B::C:D$  (read, " $A$  is related to  $B$  just as  $C$  is related to  $D$ "). Such a proportion asserts an equivalence between the ratios holding between the first two elements and the ratios holding between the last two elements. (Example: three is half of six just as four is half of eight.) Even in instances where we cannot verbalize what the relations are, a proportion asserts that they are equivalent ratios.

The mathematical model enables us to explore another feature of analogy, based upon equivalence between ratios. By alternation, we can derive from  $A:B::C:D$ ,  $A:C::B:D$ . By inversion, we can derive  $B:A::D:C$ ,  $C:A::D:B$ . By symmetry, we can derive  $C:D::A:B$ , etc. The operations lead us to further equivalences of ratios not explicit in the initial analogy, but necessarily implied by it. From  $3:6::4:8$  we derive  $3:4::6:8$ , the latter entailed by the former, but involving new ratios. Such derivations not only reveal more fully the fertility of analogy as a heuristic for discovery, but point to the need of a similarity among the elements of the ratios, which make such operations possible. With-

out some sense of similarity, the operation of alternation would not be possible. Such similarities are patent in arithmetic, but often mysterious in poetry, and too frequently problematic in philosophy.

Analogy, then, requires an equivalence of ratios and a similarity of terms that in the context of an inquiry invites an understanding of relations and a conception of terms not noticed before. This appraisal of the matter runs counter to the traditional argument, in which conceptions of the terms and understandings of the relations are taken as given. The terms are characterized as mental states (events) of oneself ( $Ms$ ) and of others ( $Mo$ ) and behavioral states (events) of oneself ( $Bs$ ) and of others ( $Bo$ ). From this beginning, we expect an analogy of the form,  $Ms:Bs::Mo:Bo$  (and, by alternation,  $Ms:Mo::Bs:Bo$ ). In such a form, all that remains is for us to explicate the nature of the relations (causal? occasional? logical?) and the character and extent of similarities between mental states and between behavioral states.

This account may make the traditional argument look drab and uninteresting; it does, of course, grossly misrepresent it. A. J. Ayer puts the argument in a clear and concise form: "I know that certain features of my own behavior are associated with certain experiences, and when I observe other people behaving in similar ways I am entitled to infer, by analogy, that they are having similar experiences."<sup>2</sup> So far, so good, since what we have can be derived by an inversion of ratios from where we began; i.e.,  $Bs:Ms::Bo:Mo$ . But this cannot be right, since Ayer has already claimed that "I have direct knowledge of my own experiences" and that "I cannot have direct knowledge of anyone else's experiences."<sup>3</sup> Since my statements about my direct experiences are not to be taken as equivalent in any sense to my statements about my overt behavior, and, further, since my knowledge of my own behavior is similar to my knowledge of another's behavior, my pattern of inference must be from my experience (directly known) through my behavior and the other's behavior (empirically derived), to the other's experience (neither directly known nor empirically derived, but analogically inferred). The similar behaviors must then serve as the mean terms to bridge the epistemological gap between my experience and that of another. On

<sup>1</sup> See Plato's *Timaeus* and Aristotle's *Posterior Analytics*. The Thomists have corrupted Aristotle's treatment of the matter—See my "Aristotle's Use of *Analogia*," *Aperion*, 1968.

<sup>2</sup> A. J. Ayer, "One's Knowledge of Other Minds," in his *Philosophical Essays* (New York, 1954).

<sup>3</sup> *Ibid.*



epistemological grounds, the proportion required seems to be: Ms:Bs::Bo:Mo.

This is *prima facie* absurd. Such a formula offers neither the equality of ratios nor the similarity of terms that we have found an analogy calls for. It cannot be derived from Ms:Bs::Mo:Bo by any rational means. It is a clear *disproportion* on the face of it. Yet, is not just such a disproportionality required by the very starting-points of the traditional argument? That argument begins from the thesis that Ms:Mo::Bs:Bo is not true. If it were, then the epistemological accessibility of Mo would be similar to Ms, just as the epistemological accessibility of Bo is similar to Bs. But that very thesis is contradicted directly by the terms in which the problem is posed. Yet, without it, the proportion cannot hold, since it is required by an alternation entailed from the original formulation.

Not only, then, does the traditional analogical argument for knowledge of other minds lack in heuristic value for discovery; it lacks in the very conditions for well-formedness required to do the job. If it is to be taken to be concerned with knowledge of other minds according to the conditions set forth for that matter as a problem, then the behavioral states of oneself and of others must serve as the mean terms of the "analogy" and disproportionality is revealed. If we begin with our paradigm for analogy, we can by alternation derive Ms:Mo::Bs:Bo, revealing that the knowledge of mental states of others is not more problematic than that of one's own mental states, thus revealing an inconsistency with the conditions set down for the problem. The upshot of the whole matter is that either the analogical argument is not analogical, or the problem of other minds is not a problem.

### III

Such a ludicrous conclusion seems possible only upon some basic misconstrual of the issues. Returning to the beginning, we find that the analogy was construed as an argument for *knowledge* of other minds. But isn't it rather the case that the issues at hand are ontological rather than epistemological? After all, a proportionality does seem to obtain on grounds of *existence*: my experience exists in relation to my behavioral states just as another's experience exists in relation to his behavioral states. By dealing with the issues from the standpoint of the way things are rather than from the way they are known by any one given individual, we seem to be able to sidestep that disproportionality exposed on

epistemological grounds. By knowledge of the existence of three terms of the proportion, we can infer reasonably the existence of the fourth.

Such a split between epistemological issues and ontological ones is itself a problematic move. I know of no case of arguing for other minds (either that they are known or that they exist) in which this sort of distinction is clearly made. Whether it ought to be made and can be made are clearly further issues. For the sake of argument, let us assume that it both can and should. Then our analogy is of the way things are rather than the way they are known. What sort of relations are claimed for the way things are? If the relations between mental states and behavioral states are taken to be logical ones, then the derivation of the fourth term can be construed to be a matter of conceptual analysis. The very conception of this sort of behavioral state logically implies the correlated conception of this sort of mental state. I understand others as having mental states correlated with behavioral states because I so understand myself. So construed, the issues have shifted from a knowing *that* to an understanding *as*, with the foundations for such an understanding somehow derivative from self-knowledge. This construal of the relations as logical would eliminate knowledge of other minds as a problem by shifting from knowing to understanding, and from contingency to necessity.

Such an appraisal seems inconsistent with the facts and inappropriate to the intentions of the advocates of the traditional argument. Whatever the status of the identity thesis regarding minds and brains, the correlations of the mental events to overt behavioral manifestations surely falls short of logical necessity. What seems to be presupposed in the argument is rather a similarity of causal relations, to be treated as contingent. On this interpretation, the fourth term of the analogy must be an empirical hypothesis rather than a logical necessity. But if this is the case, we are thrown back upon epistemological issues, and upon the absence of corroborating evidence. The only evidence for the existence of the fourth term is further analogical hypothesizing which adds no knowledge to the analogical formulation. By construing the relations as contingent rather than logical, we are forced back to the disproportionality we had hoped to avoid.

Basic to the disproportionality which confronts us is the claim to a "direct knowledge of my own experience," and with this claim to immediacy, a

claim to primacy, privacy, and incorrigibility for self-knowledge. Such claims turn on an ambiguity in talk about experience. Some talk is about having experience, about something happening to me. In this sense, my experience is clearly both immediate (nothing comes between me and it) and private (no one else has my experience). We also talk about experience as perception, as something we do. This doing has been construed in a number of ways, as a judgment (Kant), as an intentional act taking an object (Brentano), as an interaction of the organism with his environment (Gibson). Each construal involves an activity which objectivates what happens to me, providing a basis for coping with, overcoming, and/or using whatever happens to me. This activity is a mediator between what happens to me and any knowledge I may be said to have of it. It is only the experience that is the result of such mediation which we regard as knowledge. In this sense, my experience is clearly not immediate, need not be private (in so far as judgments are verbalized and objectifications functionally sharable), and thus not incorrigible (even the most rudimentary judgments are subject to error and the most simple objectifications may prove functionally inept).

This is yet only half of the flaw of the claim to immediate knowledge of one's own experience. The other half lies in making the experience the object of the knowledge. The roots for such an understanding lie in the sense-data accounts of experience, whether phenomenal or causal.<sup>4</sup> The understanding of sense data as the contents of experience lays a basis for treating the experience itself as what is most immediately known. But experience is the vehicle for knowledge, not its object, and the act of perceiving (in the sense either of consciousness taking an object or of an organism coming to terms with its environment) is not itself an object of its own activity. That an act does not act upon itself is the praxiological correlate to the theory of types.<sup>5</sup> Thus, it is not only the case that I have no strictly immediate knowledge of my own experience; it is the case that in any manner of immediacy, I have no knowledge of my own experience at all. Mental operations are not observed as what is known, but are rather presupposed as how it is known.

Feelings appear an exception to this appraisal

and pain serves as the perennial paradigm. "I feel a dime in my left pocket" describes a sensation which takes as its object whatever is in my left pocket and characterizes it as a dime. "I feel a pain in my left foot" cannot be so analyzed, since the pain is not the object of my feeling, but is the feeling itself. However we analyze the difference between the two locutions in logical form, two features remain the same. Both objectivate by name and by location. In at least these two respects, reports of feeling are subject to correction by oneself and by others ("You are right, doctor; it isn't in my foot; it's higher up in my ankle;" or "That's not a pain you feel; it's only a tickle"). In most cases, to be sure, reports of feelings are more immediate and less corrigible than reports of perception, but this is a difference in degree, not in kind.

On the model of proportionality, it became apparent that knowledge of other minds should be no more problematic than knowledge of one's own mind. The basis for disproportionality was founded upon claims to immediate knowledge of one's own mind. Having dispensed with that claim, the price for proportionality seems to be that we have no knowledge of minds at all. Not only do we not know minds as such (what would such knowledge be?), but we do not even know mental operations except as presuppositions for the objects of those operations. Neither oneself nor others are known at all, except as objects in the world; yet, both oneself and others are understood as persons. How do I come to understand persons as persons? And how do I come to understand an experience (or an action, a body, a friend, or anything else properly qualified by a personal possessive) as *my* experience.

#### IV

To pursue such questions as these, we must shift from concerns with knowing minds to concerns with understanding persons. Where the model of proportionality showed the former enterprise to be an unworkable one, it may yet help in the latter. Exploration by analogy requires a similarity among elements and an equation of ratios, not only in the initial formulation, but also in the forms derived by alternation, inversion, and symmetry. To make analogy an adequate vehicle for exploration, conceptions of the elements must be reconstructed and

<sup>4</sup> A good critique of the sense-data approach is A. M. Quinton's "The Problem of Perception," *Mind*, vol. 64 (1955), pp. 28-51. Notes of the relation of this issue to conceptions of the mind are found in such divergent sources as Gilbert Ryle's *Concept of Mind* (London, 1949), and Maurice Merleau-Ponty's *The Structure of Behavior* (Boston, 1967).

<sup>5</sup> See Ryle, *op. cit.*, pp. 197ff.

the character of the relations must be reexamined.

The shift from minds to persons is already a widely-practiced one, and comes from a variety of motivations. It has the clear advantage of not being loaded with the metaphysical and epistemological baggage that the mind/body and subject/object categories impose. The equation of the metaphysical mind with the epistemological subject is what lays the basis for the traditional formulations of the problem of other minds and the anomalies involved in talk of self-knowledge. Whether a concept of a person can be formulated to avoid such difficulties remains to be explored, but our conception need not be encumbered by self-defeating presuppositions at the outset.

The shift from minds to persons requires a comparable shift in the other terms of our analogy, a shift from behavior to action. "Behavior" seems as innocuous a term as we could want, and worth saving, but this is an insidious seeming. "Behavior" is a weasle word employed by psychologists to give the appearance of something empirically observable, while retaining the connotations of intentional operations. Once we recognize that all seeing is a seeing *as*, we must then recognize that the most "empirical" observations about motion are only possible in the light of some conceptual framework. The shift from behavior to action calls attention to the fact that we are talking about persons doing things, with all the attendant problems of intention and intentionality, of structure and function, of disposition and action. The relation of the action of persons to the "behavior" (still a weasle word) of rats to the motion of machines is left as an open question for further exploration. What is made explicit is that, when we understand something as a person, we understand his motions as actions.

With such a reconstruction of the elements of the analogy, we are enabled to explore the character of the relations of the analogy in a new light. One formulation invites a construal of the relations as actions confronting persons (or of persons opposing action). My sense of opposition here is not one of simply impeding action, but one in which the opposition confronted is understood as incapable of satisfactory objectification as a thing in the world, or of being a determinate tool for or impediment to my action; a surd opposition. By virtue of the irreducibility of this opposition to a thing or tool, the actor understands it as personal.

Another formulation invites a construal of the relations as persons intending actions. The sense of

the personal here is one of cooperation. If a person finds that another, whom he cannot reduce to object or tool in his world, none the less acts in such a way that the other fulfills the intentions of the first in his actions, he understands the other acting with him in the operation at hand. On the basis of such an understanding, he can note that his own actions are often not only unimpeded, but often aided, by the action of others. With this sense of cooperation, the understanding of community is established.

Such a characterization of the relations of the elements in our reconstructed analogy leaves those elements with a conceptual rather than an ontological status. This is the way that we understand them, not necessarily the way that they are. While I have formulated the concepts of the relations in terms of persons and actions, the existence of the relations is the condition for understanding persons as persons and actions as actions. While this does violence to prejudice derived from a heritage of Cartesian substances and logicistic atoms, it gives an accounting for senses of the personal often left out. Paranoic and animistic tendencies to appraise as personal what others regard as things and tools can be seen as understandings derived from senses of opposition and cooperation. Likewise, attempts to understand humans solely as objects in the world are attempts to understand them in purely impersonal terms.

With such a reconstruction of the analogy we can now return to the reconstructed questions with some sense of how they can be answered. What I know about myself and others I know as things in the world. I treat myself and others as persons because of confrontations and cooperations. This much has been built into the analogy itself. Understandings of the personal possessive are derivable in a variety of ways from that base. Senses of *my* action and *my* body derive from intra-personal cooperation; senses of *my* friend derive from inter-personal cooperation. The sense of *my* experience, taken to be so fundamental and immediate and crucial in the traditional argument, results from complicated interplays of confrontations and cooperations, of incompatibilities of this experience with my own past experiences and the experiences of others. While we forget the genesis of this sense after it has developed, we can continue to see its ramifications when our attention is called to understanding an experience as my experience. Such an understanding of our awareness of the world arises as a result of incompatibilities that require such a

construal. Our sense of privacy is not basic, but derivative from confrontations with our sense of community.

Explorations based on analogy usually carry implications and ramifications beyond initial conceptions and concerns. I shall mention just a few for this analogy, with no argument and with little illumination:

(1) It invites a correlate analogy between intra-personal and inter-personal relations. I understand myself as personal on the basis of confrontations and cooperations, just as I understand others as personal.

(2) It requires a four-term analogy to explicate conceptions of the relations, but it requires three-term analogies to explicate the genesis of the understanding of those relations ( $As:Po::As:Bs$  and  $Ps:Ao::Po:Ao$  for confrontations;  $Ao:Bs::Ao:Po$  and  $Ps:As::Po:As$  for cooperation.)

(3) It lays no basis for a claim that the personal is species specific to humans nor for a claim that persons are ontological entities.

(4) It requires reorderings of conceptual priorities that do violence to philosophical prejudices, but no violence to logical coherence and little to common

sense. Understanding persons precedes understanding objects, and understanding actions precedes understanding motions, both logically and genetically; understanding others as persons precedes understanding oneself as a person, and the communal precedes the private and the individual; actions rather than persons are conceptually primitive, since what counts as persons is determined both phenomenally and operationally by action.

Much still needs to be done, on this model or some other, toward understanding persons. Here I have only begun explorations by using analogical forms as heuristics for conceptual discovery. I hope that I have shown how the potential for such use of analogy has been obscured by attempts to employ it as an argument form, and how this employment with respect to concerns with other minds has been misformed by disproportional assumptions. I have indeed argued that the problems about knowledge of other minds are themselves ill-conceived. Once we have shifted our focus of concern from minds to persons, and from knowledge to understanding, then analogy may well prove a useful tool on the way toward understanding persons.

*University of Kentucky*

*Received January 2, 1973*

## VIII. THE TEMPORAL DIMENSION OF PERCEPTUAL EXPERIENCE: A NON-TRADITIONAL EMPIRICISM

DAVID M. JOHNSON

LET us draw a rough, intuitive distinction between the "form" and the "content" of what a person perceives, or what I shall call his perceptual experience. For example, when someone perceives an apple, we put the apple's red color which he sees, and the coolness and "resistance" of the apple which he feels, on the side of content; form, on the other hand, includes things like the round, bulgy shape he sees and the apple's smooth, yet also slightly grainy texture which he feels.<sup>1</sup> Present-day empiricists tend to equate the primary, important, and "real" in perception with its content while, giving a merely secondary, inferior place to form.

Consider, for example, the many present-day empiricists who follow more or less in the footsteps of Berkeley. Berkeleians emphasize the interdependence of the formal elements and content elements in our perceptual experience. Nevertheless, they also admit that strictly speaking a person becomes aware of some particular form in virtue of his awareness of some particular contents, but not *vice versa*. For example, through the felt "resistance" of an apple, a person becomes aware of this apple's smooth but slightly grainy surface. Yet we cannot say in a corresponding way that his awareness of the grainy surface is that which enables him to become aware of the apple's felt resistance. In the case of sight, similar assumptions lead Berkeleian empiricists to suppose that what we see in the strict sense of the word are lights and colors (a part of content) but visual shape (a part

of form) is just where one color leaves off and another begins.<sup>2</sup>

Analogous points can be made for the other senses as well. That is, according to these philosophers, what we smell primarily are odors, what we hear primarily are sounds, what we taste primarily are tastes; and we know the boundaries and arrangements of these odors, sounds, etc., in an inferior, not properly empirical sense.

One reason behind this prejudice of contemporary empiricists is that many of them take as the crucial "problem of perception" the question, "What are the immediate, i.e., non-inferred, objects of perceiving?"<sup>3</sup> Factors we tend to classify under the heading of form, e.g., shape, size, arrangement, seem somehow less "substantial," less like something which one could "find" and point to—and therefore less appropriate to mention in answer to this question—than factors like color, feel-, odor- and sound-like quality, etc., which we classify under the heading of content. Thus, one of these empiricists might reason as follows: "the perceived redness of an apple is something actually 'there' before us to be seen in itself; but the round shape of an apple, considered alone, is merely an infinitely thin surface or boundary of the color which is invisible, or at least not present and available in the same way to direct inspection."

Yet I deny that this prejudice in favor of content is justified. There is no reason for assuming, as contemporary empiricists do, that as soon as we

<sup>1</sup> I do not need to presuppose any special theory about the nature of perceptual experience for my purpose here. "What someone perceives" may be identified with physical objects and their parts which are capable of being literally red and bulgy, in agreement with direct realists; or it may be construed as sense-data, sense-impressions, etc., which can merely be "red" and "bulgy," in agreement with sense-datum philosophers; or it may be thought of in still some other way. The distinction between form and content of perceptual experience will remain the same, no matter which of these alternatives one chooses.

<sup>2</sup> This idea is connected with Berkeley's doctrine that our visual field consists of a large but finite number of points or *minima visibilia* which may change in color but never in either their relative position or their number. Cf. e.g., *A New Theory of Vision*, sect. 82. The more general point I am making is connected with his view that we cannot have ideas but only "notions" of the "relations and habitudes between things." Cf. *Principles of Human Knowledge*, sect. 142.

<sup>3</sup> Cf. e.g., D. M. Armstrong, *Perception and the Physical World* (London, 1961), p. xi. (However, in a later book, *A Materialist Theory of Mind* [London, 1968], Armstrong no longer seems to be happy with this question.)

identify the immediate objects of perception, all other philosophical questions about perception will also be answered. Instead we must recognize that the fundamental reason philosophers have been interested in perception is that perception provides us with knowledge.<sup>4</sup> From this viewpoint a more crucial question is the epistemological one of how what we perceive by means of the senses gives rise to knowledge. Now when we ask what part or aspect of our perceptual experience is important, we shall mean *important for the purpose of gaining knowledge*. This turns things around considerably. In particular, as I shall try to show, for the purpose of gaining knowledge, the primary and "real" in perceptual experience is found more on the side of form than on the side of content. Further, formal elements in what we perceive have more right than content to be identified with "the empirical."

The philosophers' puzzle of the reversed spectrum will illustrate the epistemological importance of the distinction I have in mind. Suppose I see red wherever and only where you see green and so on substituting contrary colors throughout the spectrum so that corresponding relations between the colors we see are everywhere preserved. For example, even though the spectrum I see is reversed abnormally, each color in it has one and only one contrary color, each color merges imperceptibly into the next color, and the whole spectrum is continuous, i.e., has no gaps, just like a normal spectrum. We would agree on a verbal level, since whatever you called "green" I would also call green, even though I would really be seeing the color that you call red. Would it ever be possible to discover this difference between us?

There are at least two possible ways one could attempt to answer this question. One approach is to ask, "Could I know both my perception of red and your perception of green in themselves so as to compare them and see that they are different?"<sup>5</sup> A second approach is to ask, "Would our knowledge of the external environment be changed as a result of the difference in our perception of colors, so that this difference in what we see could be detected indirectly?"

We can contrast these two approaches by saying that in the first approach we think of our perception of color as the object or "end" of knowledge;

in the second approach we think of it merely as a "means" to the knowledge of something else.

Let us now introduce the phrase "sensuous datum" or "given" and understand by this—in accordance with the above change in perspective—something given, by means of the senses, *for knowledge*. In particular, something counts as a sensuous datum only in so far as it contributes toward a person's acquiring knowledge non-inferentially by perceiving.

From this viewpoint, only the second approach to the preceding question about the reversed spectrum, where the emphasis is on the epistemological function of perception, is of interest. But it is just in this case that the reversed spectrum poses no particular problem. A reversed spectrum must be just as adequate a means of acquiring knowledge about the world as a normal spectrum would be, as shown by the fact that there is no verbal conflict between the two perceivers of the example despite their seeing different colors.<sup>6</sup> Unless everything one perceiver managed to learn by seeing one set of colors the other also managed to learn by seeing the other, complementary set, there would not be verbal agreement between them, because such a difference would soon show itself in conflicting knowledge claims. For example, such differences in knowledge, reflected by conflicting claims, is what allows us to distinguish people who are color-blind from those who are not.

There is no reason to think that knowledge of external things is changed merely by the fact that one of two equally adequate means rather than the other is employed to acquire this knowledge. Thus, I reject the idea that a reversed spectrum of perceived colors could be discovered indirectly from a corresponding difference in knowledge.

On the other hand, consider the "form" of color experience—in other words, the boundaries, order, and arrangements in which content elements occur. Formal elements of this kind seem more resistant and less intersubstitutable as means of acquiring knowledge than content elements. For example, if two people truthfully report that they see three evenly spaced spots of light in a dark room, it seems impossible that one of them should really be seeing four spots with uneven spaces between them and that his verbal agreement is

<sup>4</sup> Cf. J. L. Austin, *Sense and Sensibilia* (Oxford, 1962), p. 104.

<sup>5</sup> Cf. the solution to the problem along this first line proposed by W. Sellars in *Science, Perception and Reality* (London, 1963), pp. 337–338.

<sup>6</sup> Of course, this second approach is totally irrelevant to the "problem of the inverted spectrum" since the traditional problem concerns the logical or empirical possibility of having reversed sensory states *despite being able to make similar objective discriminations*.

the result of some systematic difference in the two men's experience.<sup>7</sup> There seems to be no "system" which would make sense of such a proposal. In other words, while my seeing red wherever you see green might be indefinitely "covered up" by corresponding changes in the content of what we both perceive, there seems no similar system of changes which could cover up a formal difference like my seeing four spots of light when you see three.<sup>8</sup>

The conclusion indicated is that formal elements—e.g., patterns and arrangements of perceived colors—deserve to be considered better, more objective sensuous data than content elements—e.g., perceived colors themselves.

From what I have said and the examples I have given up to this point, a reader might conclude that my distinction between the form and content of perceptual experience is identical with, or at least capable of being explained in terms of a Lockean distinction between primary (perceiver-independent) and secondary (perceiver-dependent) properties of objects. For example, such a reader might say, when someone perceives a pear, the shape, size, and texture that he finds this fruit to have count as part of the "form" of what he perceives because we think of these as properties which belong to the pear itself apart from anyone's perceptions of it. On the other hand, since we tend to assume that the pear's color, smell, and taste do not exist apart from someone's perceiving the pear, for that reason we count all the latter things as part of the "content" of what the person perceives.

I admit that there is an obvious analogy between the above distinction between the form and content of perceptual experience and the Lockean distinction between primary and secondary properties. However, one must not identify these two distinctions in an oversimple way, as one might if he considered only examples like the above case of the pear. One reason for caution here is that I conceive of those forms or patterns in perceptual experience

which are capable of acting as sensuous data not merely as arrangements in public or private space but, possibly also, arrangements of perceived elements over time.

Consider the following example. I find out by sniffing the air that my wife is wearing the same perfume this evening as she wore the night last week when we had dinner with friends at a restaurant. What I have available to me to learn this fact is one perceived secondary or perceiver-dependent property of the perfume—its odor—and no primary qualities. For example, I do not see the shape of any individual particles of perfume, nor do I locate such particles by touch. Nevertheless, the occurrence of this secondary quality at two different times enables me to make this judgment about the sameness of the perfume on the two occasions. For this reason, the sensuous datum in this case is part of the form of what I perceive rather than its content, despite the fact that the datum "consists" in nothing but perceived secondary qualities.

Let us suppose that what I have said up to this point is sufficient to show that although Berkeleyan philosophers construe perceived content elements as having more empirical worth than formal elements, there is reason to question the validity of this assumption. I shall now discuss the aspect of this general point which is my special concern in this paper—viz., the temporality of perceptual experience.

It is traditionally assumed that whatever acts as sensuous data—i.e., that contained in what we perceive, which enables us to learn intersubjective facts in the world around us without inference—must fall within the bounds of what philosophers call a specious present. This is because anything not entirely within such a present cannot be perceived but instead has to be at least partly remembered.<sup>9</sup>

Consider, for example, the following statement by Thomas Reid:<sup>10</sup>

<sup>7</sup> The possibility that either of the persons has miscounted or does not know the meaning of words like "three" or "evenly spaced" is not relevant here because we could easily detect mistakes of this kind from further things he did and said.

<sup>8</sup> However, what about reversed or mirror-images? Do they count as different shapes? And if different, are they intersubstitutable means to the same knowledge?

I am concerned here with the question of "more or less." I am not concerned positively to refute the view of contemporary empiricists but merely to correct their view's mistaken emphasis. I can allow the possibility that at least some formal differences could be systematically covered up by appropriate changes. All I need to insist on is the asymmetry that we feel about cases of the two kinds, viz., our impression that the number of such possible content changes is greater, and that these changes in content would be more easily and naturally effected than changes in formal elements. (I am indebted to R. J. Swartz whose commentary on an earlier version of the paper presented at meetings of the American Philosophical Association, Pacific Division, 1971, Beverly Hills, California forced me to make this clarification.)

<sup>9</sup> To use a phrase of Frank Ebersole's, the traditional view supposes that statements describing data of perception must be "insulated in time." Cf. his paper "On Seeing Things," *Philosophical Quarterly*, vol. 2 (1961), p. 289.

<sup>10</sup> *Essays on the Powers of the Human Mind*, vol. I (Edinburgh, 1819), p. 457.

... though in common language we speak with perfect propriety and truth when we say, that we see a body move, and that motion is an object of sense, yet when, as philosophers, we distinguish accurately the province of sense from that of memory, we can no more see what is past, though but a moment ago, than we can remember what is present; so that speaking philosophically, it is only by the aid of memory that we discern motion or any succession whatsoever. We see the present place of the body; we remember the successive advance it made to that place: The first can then only give us a conception of motion, when joined to the last.

Granted that memory somehow exerts an influence on perceiving, the question for a philosopher to try to answer is exactly how this takes place.

Some philosophers maintain that memory plays an essential role in perceiving since one has to identify and classify that which he sees, hears, etc., in order to learn something by means of perceiving, and such classification presupposes a comparison with something in the past. According to J. L. Austin, for example, this is true no matter how basic or simple an example of perceiving one chooses to consider. He says,

Any description of a taste or sound or smell or colour or of a feeling, involves (is) saying that it is like one or some that we have experienced before: any descriptive word is classificatory, involves recognition and in that sense memory . . .<sup>11</sup>

On a more complex level, these philosophers say, the influence of memory on perception may take the form of "perceiving as."<sup>12</sup> For example, suppose I see someone throw a ball and perceive it as a pitched curve ball in a game of baseball, or taste a spoonful of tea and perceive it as a blend of Darjeeling and Ceylonese. In such cases, obviously, that which I now perceive depends on and is influenced by certain things I perceived and learned about in the past, things which—at least in the sense of their still being able to influence me—I have not forgotten.

I suggested above that sensuous data can sometimes be arrangements or patterns extending over time, and not merely patterns in space. To give substance to this suggestion, I also need to maintain that memory plays an essential role in perceiving. But the preceding arguments and illustrations are useless for my purpose for two reasons.

(1) First, philosophers who talk in these terms

show that they think of memory as exerting an influence not through a length of time but entirely within a specious present. In other words, they think of past but remembered learning and experience as somehow determining the way perceived objects look and sound to us now; and only the latter is what we perceive in the strict sense of the word. They suppose that this takes place in much the same way as acts of King Henry VIII, although now past and gone, and therefore no longer in existence, may nevertheless influence affairs in present-day Britain through the agency of institutions and attitudes of people which *do* still exist. But all this is irrelevant to my idea that sensuous data are capable of extending through time. In effect, this conception of memory's influence on perception simply ignores the possibility I have suggested.

(2) Second, both the idea of comparing present perceptions with past ones and the idea of perceiving-as are ineffective as a challenge to the "official," Berkeleian view of contemporary empiricists because both presuppose a change in the meaning of the word "empirical."

Philosophers traditionally understand by "the empirical," something we perceive which indicates to us the properties of the objective world itself, unmixed with any beliefs, opinions, reactions, or interpretations of our own. For example, sense-datum philosophers admit that sense-data can be influenced by states of a perceiver like his stigmatisms or his jaundice. Their properties can also be determined to some extent by the perceiver's training or by what he is now trying to observe. But, according to these philosophers, one can go too far in this direction of allowing "subjective influences" on sense-data. In particular, to claim that what a person perceives may be partly determined by his *belief* in ghosts or his *opinion* that Jones is a lout or his *conjecture* that the world will end within 50 years or his *memory* of the time he was spanked as a child is to be in danger of leaving behind the solid empirical ground of "what one really sees" and falling instead into a marshland where one can no longer distinguish perceiving on one side from reasoning, remembering, and imagining on the other. When this happens, sense-data have lost any claim they might once have had to function as indicators of objective states of affairs, or what I call sensuous data. In other words, to conceive of sense-data in this way is to

<sup>11</sup> "Other Minds" in *Philosophical Papers* (Oxford, 1962), p. 60. Another example of this reasoning is found in A. D. Woozley's book, *Theory of Knowledge* (London, 1966), pp. 145-146.

<sup>12</sup> Cf. Austin, *Sense and Sensibilia*, *op. cit.*, pp. 100-101.



give up hope of being able to back up knowledge-claims empirically by means of them, because then one could never be sure he is not dreaming, speculating, or remembering rather than perceiving.

Yet we can only make sense of the two notions, (a) that we compare everything we perceive now with something else that we perceived earlier, and (b) the idea of "perceiving-as," if we abandon this time-honored conception of perceiving as a faculty different from believing, remembering, etc. Instead, we would now have to construe data of perception as indicating not only properties of the world but also in part, properties of our personal interpretations of the world added by means of reasoning and remembering.<sup>13</sup> For example, comparing my present experience of blue with a similar, remembered experience I had in the past is an operation which is intellectual in character rather than perceptual. The same thing can be said of construing someone's throwing a ball as a pitched curve ball in a baseball game.<sup>14</sup>

A person armed with a Tarskian semantical definition of truth can say that he knows what truth is, even if he is unable to give a single example of a sentence which is indisputably true. In a similar way, if what I have pointed out above is correct, traditional empiricists can reply to "seeing-as" theorists, that we can still know what an objective datum of perception is, by definition, even if it happens that we cannot point to one "pure" example of such a datum, unadulterated by the subjective influence of memory and interpretation.

If we have decided that the preceding "comparison with earlier experiences" and "perceiving-as" lines of thought are unacceptable because they arbitrarily change the meaning of "empirical,"

what arguments *will* do to show that perceiving involves memory? Perhaps our earlier idea of construing sensuous data as givens for knowledge will provide an acceptable answer.

Why do I reject the idea that sensuous data must always occur within a specious present? I do so because we can often distinguish the real from the illusory or one external state of affairs from another just by perceiving—i.e., without also inferring, reasoning, speculating, etc. But a static, instantaneous "picture"—even one influenced and interpreted by memory—is not a sufficient basis for doing this. For example, what allows me to know—not infer, but know directly by perceiving—that the apple I see is red rather than that it merely looks red in certain lights? I can know this, say, from the whole of the following: the apple's continuing to look red as I move around it rather than suddenly changing to look orange as soon as the sun is no longer in my eyes. Here is another example. By means of what in my perceptual experience do I learn that there is a lark singing outside my window? It cannot be from the individual notes in his song, since any of these notes could just as easily occur in the song of innumerable other birds or in the whistling of a steam kettle or screech of automobile tires. Thus again, the sensuous datum in this case must be the lark's song itself, i.e., a pattern of notes extending through time, rather than individual notes which stand in this pattern.<sup>15</sup>

Since temporal patterns sometimes constitute our sensuous givens, in these cases it must be by memory in conjunction with the senses rather than by the senses alone that we "take" that which is given us from the external world.<sup>16</sup> What I mean by "memory" here is not merely an influence from

<sup>13</sup> This comes close to the reasoning whereby old-fashioned absolute idealists ran together the world itself with our knowledge of the world. If we will not accept a point of this sort from Hegel, Bradley, or Bosanquet, why should we be willing to accept the same thing from Austin?

<sup>14</sup> If such things are *really* contained in what a person sees, can we not draw the line at any point we choose? For example, why not say it is possible to "see" the world as going to end within the next 50 years? This line of thought contains an interesting objection; but I shall not pursue it here.

<sup>15</sup> The latter example is like the case cited by Gestalt psychologists of the melody which is recognized as the same when played in different keys. The main difference between my position and that of Gestaltists is the following. Gestalt psychologists take the phenomenal "looks" of things as establishing that what we perceive primarily are patterns rather than individual elements. But I reject phenomenal considerations as inconclusive and instead arrive at a similar idea for an epistemological reason, viz., that often only patterns are sufficient to enable us to distinguish between different states of affairs non-inferentially.

<sup>16</sup> Many of Austin's criticisms of the sense-datum theory in *op. cit.*, *Sense and Sensibilia*, can be interpreted in terms of the idea that sense-datum theorists construe data as instantaneous; and this brings their claims into conflict with what we know from ordinary life. For example (cf. pp. 29 ff.) Austin denies, as against Ayer, that we use the word "perceive" in such a way that the bent stick which we "see" when a straight stick is partly submerged in water must also exist in addition to the straight stick itself. The point standing behind Austin's reasoning—we may say—is that what we see when the stick is half submerged in water is only *part* of the empirical basis or datum (in Ayer's terminology, "the phenomenon" or "the empirical evidence"—cf. p. 60) from which we find out about the existence of the stick perceptually. The whole datum presupposes more experience spread over time; this is why the "bent stick" experience is misleading when considered by itself.

something past on one's present datum, or an instantaneous recognition of some kind. Rather, it is a faculty which allows a perceiver to employ something as his datum which extends through a substantial time.

I claimed that a person employs a form or pattern in what he perceives, as a sensuous datum—and in particular, a temporal form or pattern—only “sometimes.” Let me conclude by explaining why I limit the claims of my proposed non-traditional empiricism in this way.

Sophisticated proponents of “theory-laden observation” allow that sense-experience can be “more or less phenomenal.”<sup>17</sup> According to them, in other words, sometimes sense experience is closer to “pure” experience, with very little influence upon it of theoretical interpretations, etc., and sometimes it is farther away since this influence is now more pronounced and important. But when we ask them for an account of the meaning of this “more or less”—in what circumstances one moves from one sort of experience to the other and what it means to do so—they are unable to answer.

On the other hand, my view allows for at least a partial interpretation of what it means to say that perceptual experience may be more or less phenomenal. When one is observing in a tentative, exploratory way—perhaps because he is unsure of himself or unfamiliar with the situation—he tends to focus on disjointed, momentary things and events and infer conclusions from this basis in approximately the way Berkeley and Hume thought that all perceptual knowledge was acquired. But later, when he learns better what he is doing or becomes

more secure and competent, or has a more specific task to accomplish, he can switch to using whole temporal patterns of things as the beginning point for his acquisition of knowledge. This I interpret as a movement from more phenomenal perceptual experience to less.

Consider the case of learning to understand a spoken foreign language. When we are still beginners we have to focus on each word when we hear it, and try to piece together the speaker's meaning from these single words as cues. But later when we know more about the language and have practiced it, we can take as our units of meaning complete phrases or even whole sentences—all of which take some time to perceive and cannot be compressed into one specious present. Analogously, a general practitioner must observe the characteristics of his patient's abnormal way of walking one at a time, or in an exploratory, Berkeleyan fashion, when his task is to establish whether the cause of this gait is neurological, a bone disorder, a muscle sprain, etc. However, once this has been determined, the specialist to whom the patient is referred for treatment will cease to observe things like the patient's foot raised five inches off the floor and instead observe his gait as a whole. At this point he will recognize and perceive the gait as, say, a symptom of Parkinson's disease or syphilis.

Thus, traditional, Berkeleyan empiricism, where one “infers” from perceptual experiences which are temporally discrete and punctiform, applies to only some of the observations that people make, not all of them. The same is true of the non-traditional empiricism which I have outlined here.<sup>18</sup>

*York University*

*Received September 15, 1972*

<sup>17</sup> Cf., e.g., N. R. Hanson, *Patterns of Discovery* (Cambridge, 1958), pp. 8 ff.

<sup>18</sup> I wish to thank C. A. Hooker, Wilfrid Sellars, Max Hocutt, Sydney Shoemaker, and Gilbert Ryle for help in writing this paper.

## BOOKS RECEIVED

- L'Archevêque, Paul, *Teilhard de Chardin* (Quebec: Les Presses de L'Université Laval, 1972), 289 pp. \$5.00.
- Barnes, Jonathan, *The Ontological Argument* (New York: St. Martin's Press, 1972), 98 pp. \$4.95.
- Katen, Thomas Ellis, *Doing Philosophy* (Englewood Cliffs: Prentice-Hall, Inc., 1973), 364 pp.
- Leclerc, Ivor (ed.), *The Philosophy of Leibniz and the Modern World* (Nashville: Vanderbilt University Press, 1973), 306 pp. \$15.00.
- Margolis, Joseph, *Knowledge and Existence* (New York: Oxford University Press, 1973), 289 pp.
- Pyun, Hae Soo, *Nature, Intelligibility, and Metaphysics* (Amsterdam: B. R. Grüner, 1973), 109 pp. \$7.50.
- Rescher, Nicholas, *The Coherence Theory of Truth* (Oxford: The Clarendon Press), 374 pp. \$17.75.
- Sanford, Charles L., *The Quest for Paradise* (Urbana: University of Illinois Press, 1961), 282 pp. \$5.00.
- Spisani, Franco, *The Meaning and Structure of Time* (Bologna: Azzoguidi, 1972), 161 pp.
- Weingartner, Rudolph H., *The Unity of the Platonic Dialogue* (New York: Bobbs-Merrill, 1973), 205 pp.
- White, Morton, *Pragmatism and the American Mind* (New York: Oxford University Press, 1973), 265 pp. \$8.95.
- Yandell, Keith E., *God, Man, and Religion* (New York: McGraw-Hill, 1973), 541 pp.

VOL. LXXXIII

# MIND

January 1974

Edited by  
DAVID HAMLYN

Performatives and Statements: D. HOLDCROFT  
Compatibilism and Freedom: R. YOUNG  
Sentience and Behaviour: R. KIRK  
On the Structure of Codes of Conduct: S. KÖRNER  
Disentangling the 'Cogito': W. E. ABRAHAM

Discussions:

Real Predicates and 'Exists': R. CAMPBELL  
Are Beliefs Necessary to Acting Rationally?: K. A. WALTON  
Mr. Gower on Conditionals: J. HARRISON  
Demythologising 'Or': B. MATHEWS JR.  
Evidence and Causes of Emotion: I. THALBERG  
'Ifs' Again: T. STROUP  
Classical Logic and Inexact Predicates: D. H. SANFORD  
On Doubting One's Intentions: D. F. GUSTAFSON  
Confirmationism: Replies to Critics: C. H. WHITLEY

Critical Notice:

*A Theory of Justice* by John Rawls: D. D. RAPHAEL

Book Reviews  
Books Received  
Notices

Published for the MIND ASSOCIATION by BASIL BLACKWELL

Single Issues: 40p

Annual Subscription: £1.25 net

## JOURNAL FOR THE THEORY OF SOCIAL BEHAVIOUR

Vol. 3 No. 2 October 1973

Edited by P. F. SECORD and T. MISCHEL  
Assisted by R. HARRÉ

**Contents**

|                  |   |
|------------------|---|
| PETER G. OSSORIO | Never Smile at a Crocodile  |
| JOHN SHOTTER     | Acquired Powers: the Transformation of Natural into Personal Powers           |
| HARWOOD FISHER   | Logic and Language in Defences  |
| RICHARD TOTMAN   | An approach to Cognitive Dissonance Theory in Terms of Ordinary Language      |
| MARCELLO TRUZZI  | The Problem of Relevance between Orientations for Cognitive Dissonance Theory |
| IRWIN ALTMAN     | Reciprocity of Interpersonal Exchange   |

Annual Subscription (two issues) £3.50 (\$10.50)

Single Issues £1.80 (\$5.40)

Published by BASIL BLACKWELL  
Orders to 108 Cowley Road, Oxford OX4 1JF, England

# The Philosophical Forum

A QUARTERLY

ANNOUNCES TWO SPECIAL ISSUES:

*Volume IV, No. 1*

*Devoted to Islamic Philosophy*

|                     |  |
|---------------------|--|
| MUSHIN MAHDI        | Alfarabi on Philosophy and Religion  |
| LENN EVAN GOODMAN   | Rāzī's Psychology  |
| PARFIZ MOREWEDGE    | Ibn Sina's Concept of the Self   |
| GEORGE F. HOURANI   | Ibn Sinā on Necessary and Possible Existence   |
| ALFRED L. IVRY      | Towards a Unified View of Averroes' Philosophy   |
| HENRY CORBIN        | For the Concept of Irano-Islamic Philosophy  |
| TOSHIHIKO IZUTSU    | Creation and the Timeless Order of Things: A Study in the Mystical Philosophy of 'Ayn Al-Qudāt |
| FAZLUR RAHMAN       | Mullā Sadrā's Theory of Knowledge  |
| SEYYED HOSSEYN NASR | Mullā Sadrā and the Doctrine of the Unity of Being   |

*Volume III, Nos. 3-4*

*Devoted to Georg Lukács*

Containing two articles by Georg Lukács previous unpublished in English entitled "On the Phenomenology of the Creative Process", and "On Poverty of Spirit: A Conversation and a Letter", as well as major articles by Ernst Fischer, Adam Schaff, Andrew Arato, Paul Breines, Andrew Feenberg, Willis Truitt, and Lee Baxandall.

\$15.00/yr. for Institutional subscribers; \$10.00/yr. for Individual subscribers; \$6.00/yr. for Student subscribers; \$3.50 for special Lukács Issue; \$2.75 for Single Copies, payable to: THE PHILOSOPHICAL FORUM, Department of Philosophy, BOSTON UNIVERSITY, 232 Bay State Road, Boston, Massachusetts 02215.

## RATIO

Edited by Stephan Körner

Volume XV

December 1973

No. 2

|                     |  |
|---------------------|--|
| GERHARD FREY        | Methodological Problems of Inter-disciplinary Discussions  |
| JOSEPH AGASSI       | When Should We Ignore Evidence in Favour of a Hypothesis?  |
| GRETE HENRY         | The Significance of Behaviour Study for the Critique of Reason   |
| PETER KRAUSSER      | A Cybernetic Systemstheoretical Approach to Rational Understanding and Explanation, Especially of Scientific Revolutions with Radical Meaning Change |
| FRANZ VON KUTSCHERA | An Open Problem in the Subjectivist Theory of Probability  |
| KAI NIELSEN         | Some Puzzles about Formulating Utilitarianism  |
| D. E. COOPER        | Synonymy   |
| HUNG HIN-CHUNG      | Truth by Convention  |
| C. J. F. WILLIAMS   | Prior and Ontology   |
| D. HOLDCROFT        | Escaping Taylor's Fate   |
| WILLIAM G. LYCAN    | Inverted Spectrum  |
| P. T. SAGAL         | Predicates, Concepts and Ontological Neutrality in Lorenzen  |
| TERENCE PAUL SMITH  | On the Applicability of a Criterion of Change  |
|                     | Reviews by W. L. F. GOMBOCZ, NICHOLAS NATHAN, RICHARD SWINBURNE, WILHELM TEICHNER, RAINER SPECHT   |

Price £2.25 net

Annual Subscription £4.00 net

BASIL BLACKWELL  
Oxford, England

# **STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE**

**AUGUST, 1973**

**A. F. Chalmers** Maxwell's Methodology and his application of it to Electromagnetism

**Larry Wright** The Astronomy of Eudoxus: Geometry or Physics?

**David Bloor** Wittgenstein and Mannheim on the Sociology of Mathematics

Essay Review: Objective Symmetries, Subjectivist Interpretations and Other Reflections  
(**Peter Kirschenmann**)

Edited by **Gerd Buchdahl** and **L. L. Laudan**  
Published quarterly by Macmillan Journals Limited

---

Annual Subscriptions to all countries £6.50, except US and Canada £7.50. Subscription orders should be sent, with the correct payment to Macmillan Journals Limited, Subscription Department, Brunel Road, Basingstoke, Hampshire RG21 2XS, England.

# **Australasian Journal of Philosophy**

■

**EDITOR: ROBERT BROWN**

**REVIEW EDITOR: R. S. WALTERS**

Published in May, August and December

■

Subscriptions are \$8 (Aust.) for institutions and \$5 (Aust.) for individuals  
They should be sent to the Business Manager at the  
*Department of Philosophy, School of General Studies, A.N.U.,  
Canberra, A.C.T., 2600.*

Manuscripts should be sent to the Editor at the  
*Department of Philosophy,  
Research School of Social Sciences, A.N.U.,  
Canberra, A.C.T., 2600.*

**XV WORLD CONGRESS OF PHILOSOPHY  
VARNA 1973**

**ACTS IN SIX VOLUMES-**

**IMPORTANT PHILOSOPHIC PUBLICATION  
OF THE PRESENT DAY**

In the first two volumes of the Acts are published the reports of the plenary sessions and the plenary oral examinations of the Congress. Here are included such fundamental problems as "Philosophy and Science"; "Moral and Culture"; "Technics and Man"; "Intellect and Action in Transforming the World"; "The Philosophy in the Process of the Scientific-Technical Revolution", etc. Amongst the authors of the materials are found the best philosophers of the modern world.

In the remaining four Volumes the readers will find reports in unusual interesting themes which will be discussed in the so called Problematic and Research Groups. Universal attention attract the themes: "Essence and Existence of Man"; "The Society and Its Norms, the Man and the Freedom of Personality"; "Linguistics and Antropology"; "Contemporary Historical-Philosophical Researches"; "Philosophy of Educations and Learning"; "Philosophy of Right and Political Philosophy", etc.

The Acts of the XV World Congress of Philosophy reflect the latest tendencies and directions of modern philosophy thoughts. All this makes them necessary, interesting and useful for the widest circle of readers and first of all for the workers of science and culture.

The first two Volumes will come out of print till 31st August, 1973, and the remaining four till 30th September, 1974.

Size: 16.8/23.8 cm

Soft cover

Volume: 600/700 pages

---

Subscription sum for the complete set (six volumes) US \$ 100.00 payable in advance as per Currency Account 501246.- 3 Bulgarian Foreign Trade Bank, Sofia. (Calculation Account 214241 - 200 Bulgarian Foreign Trade Bank).

To commercial enterprises agreed in advance discount.

The subscription is entered by Foreign Trade Company HEMUS, 6 Blvd. Russki 6, Sofia - Bulgaria.

# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich  
Alan R. Anderson  
Kurt Baier  
Stephen F. Barker  
Monroe Beardsley  
Nuel D. Belnap, Jr.  
Myles Brand  
Robert Butts  
Roderick M. Chisholm  
L. Jonathan Cohen

James Collins  
Anthony Cua  
Arthur C. Danto  
Richard M. Gale  
Peter Thomas Geach  
Marjorie Grene  
Adolf Grünbaum  
R. M. Hare  
Rom Harré  
Carl G. Hempel

Raymond Klibansky  
Hugues Leblanc  
Gerald J. Massey  
George Mavrodes  
Richard Rorty  
George A. Schrader  
Wilfrid Sellars  
Ernest Sosa  
Richard Taylor  
John W. Yolton

VOLUME 11/NUMBER 2

APRIL, 1974

## CONTENTS

- |  |     |  |     |
|--|-----|--|-----|
| I. PROFESSOR KEITH LEHRER: <i>Truth, Evidence, and Inference</i> . . . . .                   | 79  | IV. PROFESSOR ERNEST SOSA: <i>The Concept of Knowledge. How Do You Know?</i> . . . . . | 113 |
| II. PROFESSOR J. KELLENBERGER: <i>God and Mystery</i> . . . . .                              | 93  | V. PROFESSOR WARREN S. QUINN: <i>Theories of Intrinsic Value</i> . . . . .             | 123 |
| III. PROFESSOR DAN CRAWFORD: <i>Bergmann on Perceiving, Sensing, and Appearing</i> . . . . . | 103 | VI. PROFESSOR GARETH MATTHEWS: <i>Paradoxical Statements</i> . . . . .                 | 133 |
|  |     | VII. PROFESSOR CHARLES RIPLEY: <i>A Theory of Volition</i> . . . . .                   | 141 |



# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$14.00, increasing by \$1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$11.20, increasing by \$1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of \$6 per group of 50.

## SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers (including the Annual Monograph) and eighteen dollars for institutions (excluding the Annual Monograph which is obtainable on payment of another seven dollars). Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and five dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

## MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



## 1. TRUTH, EVIDENCE, AND INFERENCE

KEITH LEHRER

IN approaching the questions of what sentences it is legitimate to count as evidence and of what sentences it is legitimate to infer inductively from the evidence, it is essential to keep two objectives constantly before us. The first is to have as many true sentences among those we count as evidence and infer from our evidence as we can. We may put this by saying that we should seek to obtain truth. The second is to eschew falsehood whenever we can. We may put this by saying that we should aim to avoid error.<sup>1</sup> We aim at the whole truth and nothing but the truth. But what basis do we have for achieving this pair of objectives? And what sort of policy should we employ to achieve them? It is the purpose of this paper to answer these questions.

### I

*Truth and Error.* It is important to notice immediately a certain stress between the objectives mentioned. We may, if we are only interested in counting everything that is true as evidence, easily achieve our objective by simply counting every sentence as evidence; that is, for every pair of contradictory sentences count both of them as evidence. By so doing, we would immediately obtain the benefit, epistemically speaking, of counting every true statement as evidence. What makes this policy absurd is that we would also be counting every false statement as evidence, and, moreover, be contradicting ourselves at every possible opportunity. In short, we would be entirely neglecting to give adequate consideration to the second objective, the avoidance of error. Of course, we could also fully achieve our second objective at the expense of the first. For, we could avoid error simply by refusing to count any sentence as evidence unless it was beyond the risk of error. The simplest way to accomplish this would be to count nothing whatever as evidence. We would then certainly avoid error. What makes this extremely sceptical position

unsatisfactory is the complete neglect of our first objective.

The proper conclusion to draw from the preceding reflections is that a man aiming at both objectives cannot afford to sacrifice one for the certain attainment of the other without falling into the excesses of complete scepticism or universal acceptance. Instead, we must risk the certainty of securing one objective for the benefit of giving proper emphasis to both. In this observation alone we find the answer to such sceptics as Hume. For, if one is only interested in avoiding error, then no inductive inference is warranted. In any inductive inference, there is some risk of error even if the premisses of the induction are all true. On the other hand, if one is seeking to inductively infer as many true hypotheses from the evidence as one can, while at the same time seeking to avoid error, then one must consider some risk of error worthwhile. That we have no demonstration inductive inference will avoid error is not an argument against inductive inference if one attaches sufficient importance to the induction of true hypotheses. However, this reply, for whatever relief it offers from the clutches of scepticism, is only a scant beginning. For now we must turn to the very much more difficult problem of explaining how one may pay fair attention to both the objectives of seeking truth and avoiding error in what one selects as evidence and infers from it.

*Evidence and Probability.* Rather than assuming that we have evidence, we shall ask how a man can reasonably seek to accept true sentences as evidence while at the same time striving to avoid accepting any false statement as evidence. This is a difficult way to begin. It is much easier to simply postulate the existence of evidence, by appeal to perception or something of the sort, and then proceed to ask what it is reasonable to infer inductively from that evidence. But, to take a line from Russell, such postulation has all the benefits of theft

<sup>1</sup> William James noted this distinction. He said, "We must know the truth: and we must avoid error—these are our first and great commandments as would-be knowers; but they are not two ways of stating an identical commandment. . . . By choosing between them we may end by coloring differently our whole intellectual life. . . ." *The Will to Believe and Other Essays in Popular Philosophy* (New York, 1911), pp. 17–19. Quoted by R. M. Chisholm in *Theory of Knowledge* (Englewood Cliffs, 1966), p. 22, fn. 19. Also, see Isaac Levi, *Gambling with Truth* (New York, 1967), pp. 56–58.

over honest toil. We shall prefer honest toil over theft at the expense of creating a rather complicated philosophical task.

Our selection of evidence cannot be warranted by any simple appeal to experience, because experience by itself is mute. In the face of experience, we must still ask the question of what it is reasonable for us to count as empirical evidence. To what may we appeal for an answer? Experience, reason, and the testimony of others lead us to form some estimate of how good a chance a sentence has of being true, of the probability of the truth of the sentence. Now, such estimates are subjective. We may not assume that such probabilities have been objectively established, because they could only be so established on the basis of evidence. And we must not assume the existence of evidence as the basis for the very selection of evidence itself. Consequently, we must construe the estimates or probabilities on which the selection of evidence is based as being entirely subjective. They represent the convictions of a man concerning the chances of truth of a sentence in the light of his experience, reason, and what others have told him. The result is no more or less dependable than the man himself. We now turn to the question of how a man can reasonably select statements as evidence on the basis of such subjective probabilities when seeking to obtain truth and avoid error.

The subjective probabilities that must serve as the basis of evidence selection will, for our purposes, be construed as comparative probabilities. All we require as the basis of evidence selection is the capacity of the subject to compare the chances two sentences have for being true. Such a comparison can be elicited in the following way. If a man wishes to determine which of two sentences  $H$  and  $K$  has the greater subjective probability of truth for him, let him imagine himself confronted with someone giving him a fixed sum of money on the condition that he immediately bet either on the truth of  $H$  or on the truth of  $K$ . Moreover, let the man imagine that the person giving him the money is a superbookie, one who has the ability to determine immediately the truth or falsity of the two sentences. Finally, let the man imagine that the odds offered him for his bet on the truth of either sentence are the same, more specifically the odds are even. Now let us also ask the man which bet he would prefer if his only reason for preferring one bet to the other were his desire to win rather than lose the bet. In this situation, if he decides that he would prefer betting on one rather than the other,

the one on which he would bet has a higher subjective probability for him. If he decides that he would have no preference, then neither has a higher subjective probability for him than the other.

The justification of the preceding methodology is that, in the situation envisaged, the gain or loss is the same whether he wins or loses betting on  $H$  or  $K$ . Hence if he would prefer one bet to the other, it must be because, in his estimation, he has a better chance of winning by making that bet. Since he wins if and only if the sentence he bets on is true, his preferring a bet on one sentence shows he estimates that sentence to have a better chance of being true than the other.

This argument can be formulated mathematically by appeal to the theory of maximizing expected utility. The expected utility of betting on  $H$ ,  $e(BH)$ , equals the utility or value he attaches to winning a bet on  $H$ ,  $u(WH)$ , times the probability of winning, which is the probability of  $H$ ,  $p(H)$ , plus the disutility or disvalue of losing a bet on  $H$ ,  $u(LH)$ , times the probability of losing, which is the probability of  $\neg H$ ,  $p(\neg H)$ . Similarly for  $K$ , and thus we obtain the following equations:

$$e(BH) = u(WH)p(H) + u(LH)p(\neg H)$$

and

$$e(BK) = u(WK)p(K) + u(LK)p(\neg K).$$

In the situation we imagined, the value of winning a bet on  $H$  was the same as the value of winning a bet on  $K$ , and the same equality holds for the disvalues. Hence we have the following equalities:

$$u(WH) = u(WK)$$

and

$$u(LH) = u(LK).$$

Since it is a theorem concerning probability that  $p(H) + p(\neg H) = 1$ , and the same for  $K$ , it follows from the equations cited that  $e(BH)$  is greater than  $e(BK)$  if and only if  $p(H)$  is greater than  $p(K)$  and vice versa. Hence if the betting preference is rational in the sense that one bet is preferred to another in this situation if and only if it has a greater expected utility, then it follows that the subjective probability of the sentence of the preferred bet is higher for the man in question.

*A Rule of Evidence.* We shall now consider the question of how to build a rule for selecting evidence on the basis of such subjective probabilities. In what we select as evidence, it is most important that we avoid error, because if what we count as evidence is erroneous, then inference from such

evidence will be of little value in reaching true conclusions. Hence, it is important that we require that the sentences we count as evidence have a very strong chance of being true. We must set the mark for evidence selection high, so that, comparatively speaking, sentences counted as evidence will be ones that have a high subjective probability of being true rather than false. We now turn to articulation of a rule for the selection of evidence which meets this requirement while conforming to our subjective starting point.

We assume that we can compare the subjective probability of two sentences that compete with each other for the status of evidence. We can determine whether or not the subjective probability of one sentence is greater than that of a second. For a sentence to be selected as evidence, it must have a higher subjective probability of being true than those with which it competes. We have already illustrated how one may compare the subjective probability of two sentences without assigning a quantitative value to the probabilities. Hence, though we shall employ a probability functor when we say that the probability of a  $H$ , symbolically,  $p(H)$ , is greater than the probability of some sentence with which it competes, we need not assume that any single quantitative assignment of values is correct, or, for that matter, that we can even find any plausible method for assigning such quantitative values. The theory of subjective probability as developed to date warrants the conclusion that there are suitable methods for assigning quantitative subjective probabilities to sentences, but we shall not assume that any single assignment is correct.<sup>2</sup> Any two assignments of numerical values that yield the same comparative ordering of the subjective probabilities of sentences result in the same selection of sentences as evidence. Any quantitative assignment yielding the comparative ordering of the subjective probabilities of sentences obtained by the method sketched above is as satisfactory as any other yielding that ordering.

Moreover, it is not at all difficult to formulate a rule for the reasonable acceptance of sentences as evidence. The rule is as follows: accept a sentence as evidence if and only if it has a higher subjective probability than any sentence with which it competes for the status of evidence. This rule is, of course, in need of more precise articulation. Having explicated the notion of comparative subjective

probability, the task that remains is to explain what a sentence must compete with for the status of evidence.

*Competition and Conflict.* It has been common for philosophers to think of evidence as that which is certain. Hence, it might seem tempting to consider only those sentences as evidence which have a subjective probability of unity, that is, are estimated to have no chance whatever of being false. Such an assignment would fit with our objective of avoiding error but at considerable epistemic expense. Some are inclined to consider only tautologies as having no chance of being false, all other sentences being exposed to at least some risk or error. If such a man were to restrict his evidence to what was certain he would have nothing beyond tautologies as evidence. This would surely be overly restrictive. We should, instead, be willing to risk some chance of error for the benefit of increasing the chances of adding true empirical sentences to our body of evidence. Accordingly, we shall only require that an empirical sentence compete favorably with other empirical sentences for the status of evidence. The problem is to say with which empirical sentences a given sentence must compete.

We cannot, of course, appeal to any evidence to decide which sentences compete with which or we shall, once again, find ourselves reasoning in a circle. We shall be assuming some sentences are to be counted as evidence in order to decide which sentences may be counted as evidence. We must not follow that course. Without appealing to evidence, we can only appeal to logical relations which we discern to hold between sentences. Continuing to search for the light of truth without tripping into the dark pit of error, we must assume that a sentence competes with every sentence that, as a matter of logical possibility, it could conflict with. One sentence conflicts with a second when the truth of the first insures the falsity of the second. Hence, provided it is logically possible that if one sentence is true, then a second is false, such sentences could conflict, and, therefore, we assume that they compete for the status of evidence. To make any other assumption would be to assume, as a matter of evidence, that the two sentences did not conflict, and, as we have noted, such an assumption would be quite illegitimate.

The problem remaining to be solved before we can explain when one sentence competes with

<sup>2</sup> For material on subjective probabilities see, *Studies in Subjective Probability*, ed. by Henry Kyburg, Jr. and Howard Smokler (New York, 1964) and Richard Jeffrey, *The Logic of Decision* (New York, 1965). For application to evidence selection, see my "Induction and Conceptual Change," *Synthese*, vol. 23 (1971), pp. 206-225.

another for the status of evidence is to determine when it is logically possible for one sentence to compete with a second in the sense that if the first is true, then the second is false. The question, thus posed, admits of a ready answer. For it is logically possible for one sentence to conflict with a second in the requisite sense if and only if the first does not logically imply the second. If one sentence logically implies a second, then, if the first is true, the second must be true also. On the other hand, if the first does not logically imply the second, then it is logically possible that if the first is true, then the second is false. Hence, we reach the conclusion that one sentence competes with a second if and only if the first does not logically imply the second.

*Evidence and Hypothesis.* However, this formulation needs some further qualification. For, it is entirely unrealistic to require that a man consider all those sentences not logically implied by some sentential candidate for the status of evidence before awarding the candidate the office. There will be ever so many sentences not logically implied by a sentence whose consideration is irrelevant to the question of whether it should be accepted as evidence. We must add some condition of relevance. Of course, again we must accomplish this without surreptitiously assuming that some sentences already have attained the status of evidence. We can accomplish this without departing from the subjective way, by noting that the question of whether a sentence is to be accepted as evidence always arises in some context where we are asking whether the sentence is to be accepted as evidence *for or against some hypothesis*. The question of whether any given sentence is evidence raises the further question—evidence for what? We require some hypothesis for which, or against which, a sentence is to be considered as evidence in deciding whether to count it as evidence.

We refer to the hypothesis evoking the question of what sentences are to be counted as evidence for or against it as the *touchstone hypothesis*. It is by coming into epistemic contact with such a hypothesis that a sentence becomes an active candidate for the status of evidence. The sentences that compete for the status of evidence in such a context are the ones that are relevant to the touchstone hypothesis. The conception of relevance must be construed subjectively here. And it is quite simple to say what it means for a sentence to be relevant to another in terms of comparative subjective probability. A sentence  $S$  is relevant to a hypothesis  $H$  if and only if the probability of  $H$ ,  $p(H)$ , is either less than the probability of  $H$  on  $S$ ,  $p(H, S)$ , in which case

$S$  is favorably relevant to  $H$ , or  $p(H)$  is greater than  $p(H, S)$ , in which case  $S$  is unfavorably relevant to  $H$ . Thus, with respect to a touchstone hypotheses  $H$ , a sentence  $S$  competes for the status of evidence only if  $S$  is relevant to  $H$ .

However, this proposal still requires some modification to make the set of competing sentences manageable. We can obtain some refinement of this set by constructing a *partition* from the set of relevant sentences. Let us assume that the set of sentences thus considered is finite, and, therefore, that the members can be placed in numerical order. Hence we have the sentences  $S_1, S_2$ , and so forth to  $S_n$  all of which are relevant to the touchstone hypothesis  $H$ . We obtain a partition by constructing maximal conjunctions containing every  $S_i$  or its negation but not both in numerical order. We then exclude those maximal conjunctions that are self-contradictory. The resulting set of maximal conjunctions will be logically inconsistent in pairs. It is logically impossible that any two of them could both be true; and the set is logically exhaustive: it is logically necessary that at least one of them should be true. Moreover, each of the original sentences  $S_i$  will be logically equivalent to some disjunction of members of this partition. Thus, for simplicity we can restrict competition to members of the partition and disjunctions of such members. We shall call a set consisting of members of such a partition and all disjunctions of such members, with disjuncts occurring in numerical order, a *competition set* for  $H$ . Actually, some members of a competition set for a touchstone hypothesis  $H$  will contain members that are not relevant to  $H$ . However, since these sentences are constructed out of sentences that are relevant by conjunction, we shall require that other sentences of the competition set compete with them for the status of evidence.

*Evidence and Competition.* We are now in a position to present our final explication of competition. First, our notion of competition is made relative to a competition set for some hypothesis  $H$ . We thus obtain the following definition of competition for members of the same competition set. Two sentences belonging to a competition set for  $H$  compete with each other if and only if either the two sentences logically conflict, that is, are logically inconsistent with each other, or the two sentences could conflict, that is, neither logically implies the other. With this definition of competition, we proceed to formulate a rule for the acceptance of sentences as evidence relevant to some hypothesis. The rule is as follows:

- (RE) Accept  $E$  as evidence relevant to  $H$  if and only if  $E$  is a member of a competition set  $C$  for  $H$  and  $p(E)$  is greater than  $p(S)$  for any  $S$  in  $C$  with which  $E$  competes.

If we expand (RE) in terms of the definitions offered above, it says to accept  $E$  as evidence relevant to a touchstone hypothesis  $H$  if and only if  $E$  is a member of a competition set  $C$  for  $H$  and the probability of  $E$  is greater than the probability of any other member of the competition set which is either logically inconsistent with  $E$  or which neither logically implies nor is logically implied by  $E$ . Having formulated the rule, let us now examine the rationale for doing so.

First, consider what sentences in a competition set for a hypothesis the rule tells us to accept as evidence. We can do this by reference to the partition in terms of which the competition set was formed. A member of the partition is accepted as evidence by the rule if and only if it is more probable than its denial. A disjunction of members of the partition is accepted as evidence if and only if it is more probable than any disjunction of members which it does not logically imply and which is not logically implied by it. This means that a disjunction of members is accepted as evidence if and only if it is more probable than the denial of its least probable disjunct. These results guarantee that if a sentence is accepted as evidence and that sentence is true, then it is more probable than any false sentence belonging to the same competition set. This result is desirable if the truth of sentences accepted as evidence is to be explained by the subjective probabilities. If a sentence accepted as evidence could be true when some equally probable sentence of the same competition set were false, then it would be implausible to contend that the truth of the sentence accepted as evidence was explained by the subjective probabilities.

There are other features of (RE) that favor it. First, the rule is conjunctive. If two sentences,  $D$  and  $E$  belonging to a competition set for  $H$  are each accepted as evidence by (RE), then so is the conjunction of those two sentences. Moreover, there is some sentence  $T$  accepted as evidence by (RE) from such a competition set that logically implies

every other sentence of the set accepted as evidence by (RE). This sentence  $T$  is the total evidence accepted by (RE). Finally,  $T$  is logically consistent. Such considerations, though they add to the plausibility of (RE), fall short of constituting a substantive argument in defense of the principle. We must find some more systematic justification.

*Evidence and Epistemic Utility.* We can provide an argument for (RE) by appeal to the theory of expected utility adopted to epistemological choices.<sup>3</sup> Let us consider a man faced with the decision of whether to accept some sentence as evidence relevant to some touchstone hypothesis. There are many consequences that might result from his accepting the sentence. But we are here interested only in those consequences that are relevant to the epistemic objectives of seeking truth and avoiding error. This greatly simplifies the problem. For, now there are but two consequences to consider. The first is that the sentence accepted as evidence is true, in which case the man obtains the truth he seeks, and the second is that sentence is false, in which case he falls into error. The expected utility of accepting a sentence as evidence will be a function of the value or utility a person attaches to each of these outcomes and to the probability of each outcome. More specifically, the expected utility of accepting  $E$  as evidence,  $e(AE)$ , will then be equal to the probability of  $E$  being true,  $p(E)$ , times the utility of accepting  $E$  when  $E$  is true,  $u(TE)$ , plus the probability of  $E$  being false,  $p(-E)$ , times the utility of accepting  $E$  when  $E$  is false,  $u(FE)$ . The formula is as follows:

$$e(AE) = p(E)u(TE) + p(-E)u(FE).$$

To offer a decision-theoretic justification of (RE) in terms of the theory of expected utility we must show that the results we obtain from RE are warranted by the policy of seeking a maximum of expected utility as specified in this equation. This leaves us with two problems. The first, and least difficult, is that the equation appears to commit us to a quantitative assignment of subjective probabilities and we have vowed to assume nothing more than comparative probabilities. Secondly, it also appears that we must find some quantitative measure of the two utilities in the equation. We shall undertake a

<sup>3</sup> The employment of *epistemic* utility functions is due to Carl G. Hempel, "Deductive-Nomological vs. Statistical Explanation," in Herbert Feigl and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Vol. III (Minneapolis, 1962), pp. 98-169. His use of such functions was restricted to the acceptance of sentences as hypotheses. Others who followed him include Isaac Levi in *Gambling*, *op. cit.*; Jaakko Hintikka and Juhani Pietarinen, "Semantic Information and Inductive Logic," in *Aspects of Inductive Logic*, ed. by Jaakko Hintikka and Patrick Suppes (Amsterdam, 1966), pp. 96-112; Risto Hilpinen, *Rules of Acceptance and Inductive Logic*, *Acta Philosophica Fennica*, Fasc. 22 (1968).

solution of the second problem first by reducing the utilities in question to probabilities. We shall then solve the first problem by showing that any probability function consistent with our comparative probabilities will suffice for the purpose of justifying (RE). We do not require any method for deciding between such functions, because any arbitrarily selected function consistent with our comparative probabilities will be as good as any other.

*Probability and Utility.* We remarked earlier that the avoidance of error is paramount in accepting sentences as evidence. Hence the disutility of accepting a false sentence as evidence,  $u(FE)$ , will be high for most candidates for the status of evidence. How should this value be determined? The answer to this question is to be found by asking what number adequately expresses the loss we suffer if we accept  $E$  as evidence relevant to some touchstone hypothesis when  $E$  is false. A suitable measure of this loss is the probability of the strongest competitor  $E$  has within the competition set for  $H$ . What we lose when we accept a false sentence as evidence is just the chance of truth, the probability of truth, that we have foregone by accepting  $E$  as evidence instead of some strongest competitor. Since competition is determined in terms of probability, the most probable being the winner in the competition, a strongest competitor of a sentence for the status of evidence is a most probable competitor.

With this insight to guide us, and letting  $E^*$  be any strongest competitor of  $E$ , in case there is more than one equally probable strongest competitor, we accept the following equation:

$$u(FE) = -p(E^*).$$

The utility of accepting a sentence as evidence when it is true is a gain that may be measured in terms of avoiding a loss. If the maximum gain is unity, as it is when we measure utilities in terms of probabilities, then the gain of accepting what is true is equal to that maximum of unity minus what we would have lost had we instead been in error. Thus we obtain the following equation:

$$u(TE) = 1 - p(E^*).$$

These equations reflect a policy of seeking to avoid error by measuring the loss of accepting a false sentence as evidence in terms of the chance for truth we have passed by in accepting a strongest competitor instead, and measuring the gain of accepting a true sentence as evidence in terms of what remains when we subtract the loss that we

would have suffered by being in error from the amount of a maximum gain.

These values can be obtained once we have a method for finding a most probable competitor for a statement  $E$  in a competition set for a hypothesis  $H$ . This is not difficult to do. The most probable competitor for any member of the defining partition of the competition set is a disjunction of all the other members of the partition. A most probable competitor for any contingent disjunction of members in the competition set can be constructed by finding any least probable disjunct and then forming a disjunction of all the other members of the partition except that one. This disjunction will be a strongest competitor in the competition set. These results follow directly from the principle of the additivity of probability, that is, the principle that the probability of a disjunction of mutually inconsistent sentences, members of the partition for example, is equal to the sum of the probabilities of the sentences disjoined.

By substituting in our equation for expected utility we obtain the following equation:

$$e(AE) = p(E)[1 - p(E^*)] + p(-E)[-p(E^*)].$$

By substituting in terms of the equation

$$p(-E) = 1 - p(E),$$

which follows from the calculus of probability, and performing some algebraic operations we can reduce our equation to the following:

$$e(AE) = p(E) - p(E^*).$$

This tells us that the expected utility of accepting  $E$  as evidence is equal to the difference between the probability of  $E$  and the probability of the strongest competitor of  $E$ .

Now if we follow the policy of accepting as evidence that sentence with a maximum of expected utility, we shall accept the sentence whose probability exceeds that of its competitor by the largest amount. We have no reason to doubt that it would be reasonable to accept such a sentence as evidence. However, we need not restrict ourselves to the acceptance of just that single sentence as evidence, and, indeed, by so doing we pass up the opportunity of obtaining as much expected utility as we can. In the customary application of utility theory, a man is confronted with a set of alternatives which are mutually exclusive in pairs so that he can choose but one. To obtain a maximum of expected utility it is reasonable for him to choose a maximal alternative. Choosing that one, he can choose no other. But we are in a different position when

accepting sentences as evidence. We can accept more than one. Hence, if we are guided by the policy of getting all the expected value one can, then we should accept all those sentences as evidence whose acceptance has positive expected utility. By so doing we obtain all the expected utility available to us. This is the reasonable policy in terms of the theory of expected utility.

The preceding argument has, of course, lead us, by a decision theoretic route, directly back to rule (RE). For our decision theoretic policy of accepting all those sentences as evidence whose acceptance has a positive expected utility is precisely the policy of accepting those sentences as evidence which are more probable than any sentence with which they compete for the status of evidence. Those sentences whose expected utility is positive are the ones that are more probable than any strongest competitor. Of course, a sentence more probable than any strongest competitor is more probable than any competitor, and vice versa. Thus, what is reasonable to accept as evidence, when one is seeking a maximum of expected epistemic utility, is exactly what rule (RE) tells us to accept. Therefore, rule (RE) dictates the acceptance of just those statements as evidence relevant to a touchstone hypothesis that it is reasonable to accept for the purpose of seeking truth and avoiding error.

Before turning to a discussion of what sentences it is reasonable to infer as hypotheses from evidence, it is important to notice that our decision theoretic justification of (RE) does not presuppose that we have any method for choosing any single quantitative probability function as correct. Our justification is equally cogent for any probability function that preserves the comparative probabilities on which rule (RE) is based. All that we require to determine whether it is reasonable to accept a sentence as evidence is the comparative probability of that sentence to other sentences in the competition set. Thus, for our purposes, the quest for a single quantitative probability function is otiose.

## II

*Induction and Reasonable Acceptance.* We now turn to the question of when it is reasonable to infer inductively a hypothesis from the evidence we have accepted as relevant to it. To say that it is reasonable to infer inductively a hypothesis from evidence is, we shall assume, equivalent to saying that it is reasonable to accept the sentence as a hypothesis by induction from the evidence. We shall now artic-

ulate and justify a rule for the acceptance of sentences as hypotheses by induction from the relevant evidence. Our approach to the problem will exactly parallel our approach to the problem of accepting sentences as evidence.

We now ask what sentences a sentence competes with for acceptance as a hypothesis by induction from the evidence. A competition set may be constructed from a partition constructed in the manner indicated when we discussed the selection of sentences as evidence. However, the sentences we accept as evidence logically exclude some members of the original competition set. The competition is thus reduced by the evidence. Competition is now restricted to members of the original competition set that are neither inconsistent with the evidence  $E$  nor are disjunctions containing disjuncts inconsistent with  $E$ . We call such a set of sentences a *reduced* competition set. Thus, given a hypothesis  $H$ , we now can give an initial formulation of our rule of induction as follows: It is reasonable to accept a sentence as a hypothesis by induction from the evidence if and only if it competes favorably with members of a reduced competition set for that status. As in the case of evidence, we need to provide an explanation of the appropriate conception of competition.

Let us first consider the objectives relevant to the selection of a hypothesis on the basis of evidence. Our purpose, generally expressed, is to obtain truth and avoid error. We said in our discussion of the selection of sentences as evidence that overriding concern in such selection was to avoid error. For, if our evidence consists of false sentences there is little hope of arriving at true hypotheses by induction from the evidence. However, having given an upper hand to the objective of avoiding error in the acceptance of sentences as evidence, it is time to give the objective of obtaining truth the upper hand in the acceptance of sentences as hypotheses by induction. For, just as it must be our primary goal to avoid error in what we take for our base of evidence, so must it be our goal to reach out for the truth in what we hypothesize from that base.

To give an upper hand to the search for truth, we should aim at accepting sentences as hypotheses which have great informative content. The more informative a hypothesis is, the more truth we can obtain about the world by accepting it. We have no guarantee that any hypotheses we accept by induction from the evidence will be true. Anyone who requires a demonstration that inductive inference leads to the acceptance of a true hypothesis will be



inevitably lead into sophistry or Humean scepticism. To inductively infer a hypothesis from the evidence is to risk error for the sake of accepting a hypothesis which, if true, gives us information about the world. It is only if one is sufficiently motivated by the goal of obtaining truth that one will consider the risk worthwhile. We do not pretend to offer any proof that one should be willing to risk error to obtain truth. Here we arrive at the bottom rock of fundamental epistemic preferences.

*Competition and Equal Content.* If, however, we agree that our objective is to accept informative hypotheses for the purpose of increasing our information about the world, if what we accept is true, then the explication of competition becomes straightforward. If we ask what sentences in a competition set a hypothesis competes with for acceptance as a hypothesis by induction from the evidence, the answer is only those hypotheses having the same amount of informative content. Competition is thus layered in terms of the level of content. Moreover, for a sentence to compete favorably for acceptance as a hypothesis, it must be more probable on the basis of the evidence than those hypotheses with which it competes. So, it is reasonable to accept a sentence as a hypothesis by induction from the evidence if and only if it is more probable on the evidence than other sentences of a reduced competition set having the same amount of informative content. As a result of being more probable, there is less risk of error in accepting that sentence as a hypothesis than in accepting other sentences which, if true, offer us the same amount of truth about the world.

All that remains to complete the explication of the concept of competition is to provide some method for determining when two sentences in a reduced competition set have the same amount of informative content. The reduced competition set contains members of the defining partition and disjunctions of such members in numerical order. A most natural measure of equality of content is the following: all members of the partition in the reduced competition are equal in content as are all two-membered disjunctions, three-membered disjunctions, and so forth in that set.<sup>4</sup> An intuitive justification for this method of determining equality of content can be found if we take as our partition sentences describing the outcome of a one hundred

ticket lottery when our evidence tells us the lottery is fair and exactly one of the one hundred tickets has been picked. The partition consists of sentences saying that a given number is the winner. The competition set consists of disjunctions of such hypotheses. The sentence "the number one ticket is the winner" gives us the same amount of information as the sentence "the number two ticket will win." Each of these hypotheses has greater content than the sentence "either the number one ticket is the winner or the number two ticket is the winner" which has the same amount of content as the sentence "either the number three ticket is the winner or the number four ticket is the winner."

*An Inductive Rule.* Having defined equality of content, we can say that one sentence in a competition set is an *inductive competitor* of another if and only if the two sentences belong to the reduced competition set and have equal informative content. A sentence  $H$  beats a sentence  $K$  in such competition if and only if the probability of  $H$  on the evidence  $E$ ,  $p(H, E)$ , is greater than the probability of  $K$  on  $E$ ,  $p(K, E)$ . Thus we obtain the following inductive rule:

- (IR) Accept  $H$  as a hypothesis by induction from evidence  $E$  out of a reduced competition set  $C$  for  $H$  if and only if  $p(H, E)$  is greater than  $p(K, E)$  for any  $K$  in  $C$  that is an inductive competitor of  $H$ .

This rule may be incorporated in a broader principle of acceptance by including among those sentences it is reasonable to accept, not only the sentences that it is reasonable to accept as evidence and as hypotheses by induction from the evidence, but also all those sentences in the competition set that are deductive consequences of the sentences accepted as evidence and hypotheses. The reason for accepting such deductive consequences from the competition set is that they cannot possibly be false if the evidence and hypotheses are true, and, hence, we run no additional risk of error by accepting such sentences. What we can accept without any additional risk of error is, epistemically speaking, a bargain not to be ignored. Hence we propose the following principle of acceptance:

- (RA) Accept all those sentences from a competition set that are deductive consequences

<sup>4</sup> Measures of content yielding such results were proposed first by Rudolph Carnap and Yehoshua Bar-Hillel in *An Outline of a Theory of Semantic Information*, Techn. Report No. 247, M. I. T., Research Laboratory of Electronics, 1952, and reprinted in Yehoshua Bar-Hillel, *Language and Information* (Reading, Massachusetts, 1964), pp. 221-274. They were followed by Jaakko Hintikka and J. Pietarinen, in "Semantic Information," *op. cit.*, and Isaac Levi in *Gambling*, *op. cit.*

of sentences accepted as evidence by rule (RE) or as hypotheses by induction from evidence by (IR).

Having laid down these principles we now turn to the task of further justifying them.

As an initial step of justification, we may note the consequences of the rules. In terms of (IR), the results are fairly obvious. If a single member of the partition in the reduced competition set is more probable on the evidence than any other, it may be accepted as a hypothesis by induction. If a disjunction containing  $n$  members of the partition in that set is more probable on the evidence than any other such  $n$  membered disjunction, then it may be accepted as a hypothesis. By adding the deductive consequence condition embodied in (RA), we obtain the conclusion if any member of the partition in the reduced competition set is more probable on the evidence than any other, then we may accept that sentence and all disjunctions in the competition set containing that sentence as a member. If two or more members of the partition in the reduced competition set are equally probable on the evidence but more probable than other such members, then we may accept a disjunction of those most probable members and all disjunctions of members containing all those most probable members as disjuncts. The sets of sentences accepted by the three rules will be logically consistent.

Rules (IR) and (RA) tell us to accept the most informative sentence as a hypothesis by induction that we have less reason to think to be erroneous than any other equally informative hypothesis. Thus rule (IR) requires us to pay some heed to the goal of avoiding error by requiring that the hypothesis we accept have less chance of being erroneous than any other of equal content. On the other hand, if there is some hypothesis having less chance of being erroneous, however small the margin of difference, than others that are equally informative, then the rule directs us to accept that sentence as a hypothesis by induction from the evidence. For, if such a hypothesis is not erroneous, it yields as much truth about the world as any sentence we had any reason to believe not to be erroneous. In this way the rule, though favoring the search for truth, carries on the search with an honest regard for the avoidance of error.

*Induction and Expected Utility.* All of the preceding remarks remain at the level of intuitive justification. We can provide an argument for (IR) and (RA)

by appeal to a suitably restricted application of the theory of expected utility. Though there are many consequences that may result from accepting a sentence as a hypothesis by induction from the evidence, we shall again only be interested in those consequences directly relevant to seeking truth and avoiding error. Thus, we consider two possible results of accepting a sentence as a hypothesis by induction from the evidence; accepting the sentence when it is true and accepting the sentence when it is false. The expected utility of accepting  $H$  as a hypothesis by induction from  $E$ ,  $e(IH, E)$  will be a function of the utility of accepting  $H$  as a hypotheses by induction from  $E$  when  $H$  is true,  $u(TH, E)$ , and the disutility of accepting  $H$  as a hypothesis by induction from  $E$  when  $H$  is false,  $u(FH, E)$ , as well as of the probabilities of these outcomes. The equation for expected utility is then as follows:

$$e(IH, E) = p(H, E)u(TH, E) + p(-H, E)u(FH, E).$$

We shall again, for convenience, begin by supposing that we have some method for determining a single correct assignment of quantitative probabilities and subsequently show that this supposition is superfluous. The problem to which we now turn is that of finding some measure of utility and disutility.

When we accept a sentence as a hypothesis by induction, we are accepting a hypothesis which, if true, will be highly informative. This suggests that the utilities in question are some function of the informative content of  $H$  on  $E$ ,  $\text{cont}(H, E)$ , and of the informative content of the denial of  $H$  on  $E$ ,  $\text{cont}(-H, E)$ . To apply this suggestion we require some measure of content. One method for obtaining a measure of content is derived from our earlier principle that members of the defining partition in the reduced competition set have the same content, two membered disjunctions in the set have the same content and so forth. We can obtain a measure of content preserving these equalities and also reflecting the judgment that a member of the partition has greater content than a two-membered disjunction of such members, which in turn has greater content than a three-membered disjunction of such members and so forth.

Let  $n$  equal the number of members of the defining partition in the reduced competition set, and, for any sentence  $H$  in that set, let  $h$  equal the number of members of the defining partition disjoined in  $H$ , if  $H$  is a disjunction, or let  $h$  equal

one, if  $H$  is a member of the partition. We define a measure function for any hypothesis  $H$  in the reduced competition set as follows:

$$m(H, E) = h/n.$$

From this measure function, we define content.

$$\text{cont}(H, E) = 1 - m(H, E)$$

and

$$\text{cont}(-H, E) = m(H, E).^5$$

As an illustration of the values obtained, suppose that we have ten members of the partition in the reduced competition set. The content of such a member will be  $9/10$ , of a disjunction of two such members  $8/10$ , of three members  $7/10$ , and so on. The content of the denial of a member is  $1/10$ , of the denial of a disjunction of two members  $2/10$ , and of the denial of a disjunction of three members  $3/10$ , and so on.

One familiar method for defining utilities in terms of content, which is appealing for its simplicity, is to equate the utility of accepting a sentence as a hypothesis by induction when it is true with the content of the hypothesis and the disutility of accepting a sentence as a hypothesis by induction when it is false with the minus value of the content of the denial of the hypothesis. We then obtain the following two equations:

$$u(TH, E) = \text{cont}(H, E)$$

$$u(FH, E) = -\text{cont}(-H, E).$$

We shall ultimately wish to reject these equations. But by revealing the defect in this simple proposal we can illustrate the need for the more complicated modification we shall advocate.

Now, turning once again to our equation for expected utility and substituting in terms of the equations given, we obtain

$$e(IH, E) = p(H, E)(1 - m(H, E)) + p(-H, E)(-m(H, E))$$

which reduces to

$$e(IH, E) = p(H, E) - m(H, E).$$

It is important to notice that the same ranking of expected utilities may result if we multiply  $m(H, E)$  by some constant  $k$ . Hence, if it is only the ranking of the expected utility of accepting sentences as hypotheses that determines whether or not a sen-

tence is accepted, then, for the purposes of such acceptance, the equation

$$e(IH, E) = p(H, E) - m(H, E)k$$

may yield results equivalent to the equation above. This equivalence becomes important when the antecedent probabilities of the members of the partition in the reduced competition set are all the same. From the additivity of probability, it follows that for some constant value of  $k$  the following equation holds for any sentence  $H$  in the reduced competition set:

$$p(H) = m(H, E)k.$$

This is so because the measure function,  $m(H, E)$ , assigns equal values to all members of the defining partition set in the reduced competition set and is also additive, that is, the value of disjunctions of such members is equal to the sum of the value of the members.

Thus, when the members in question have equal antecedent probabilities, the equation for expected utility just given yields values equal to the following equation:

$$e(IH, E) = p(H, E) - p(H).$$

By this equation, expected utility is a measure of the relevance of  $E$  to  $H$ , that is, if the expected utility is positive, then  $E$  has increased the probability of  $H$ , if negative, then  $E$  has decreased probability of  $H$ . The equation could have been obtained from defining content in terms of antecedent probabilities and utilities in terms of those probabilities as follows:

$$\text{cont}(H, E) = 1 - p(H)$$

$$\text{cont}(-H, E) = p(H)$$

$$u(TH, E) = 1 - p(H)$$

$$u(FH, E) = -p(H).$$

These equations may yield the same ranking of expected utilities as the original equations under the condition cited above, to wit, the equal antecedent probabilities of members of the defining partition in the reduced competition set.

*A Rule of Acceptance.* Similar equations have been advocated by Levi, Hilpinen, and Hintikka and Pietarinen.<sup>6</sup> However, as both Levi and Hilpinen have pointed out, there can be more than one hypothesis with a maximal expected utility. Indeed, if  $p(H, E) = m(H, E)$  for every  $H$  in the reduced competition set, which may occur, then every

<sup>5</sup> Isaac Levi, *Gambling*, *op. cit.*, pp. 69-70.

<sup>6</sup> Isaac Levi, *Gambling*, *op. cit.*, pp. 69-71, Risto Hilpinen, *Rules*, *op. cit.*, pp. 105-119, and Jaakko Hintikka and J. Pietarinen, "Semantic Information," *op. cit.*, p. 99.

sentence will have a maximal expected utility. Hence, if we were to adopt the principle to accept sentences as hypotheses by induction that have a maximal expected utility, we would have a rule telling us to accept sentences as hypotheses which contradicted each other. Our interest in avoiding error should suffice to deter us from allowing this procedure. Neither can we subscribe to the policy of accepting all sentences as hypotheses having a positive utility, because two members of the partition which contradict each other may each have a positive expected utility.

The foregoing considerations suggest that we should require that the expected utility of a sentence we accept as a hypothesis by induction be both maximal and positive. There can be more than one sentence whose expected utility is maximal and positive. This arises from the fact that  $e(IH, E)$  as specified above is additive. Hence, if some member of the defining partition in the reduced competition set has an expected utility of zero, then the disjunction of that member to a member whose expected utility is positive and maximal will also be positive and maximal. However, the set of sentences whose expected utility is both maximal and positive will be mutually consistent. So one cannot object to the policy of accepting sentences as hypotheses that are maximal and positive on the grounds that it will lead to contradiction and error.

Others, Hilpinen and Levi, have argued that we must have a rule for breaking ties among positive and maximal sentences.<sup>7</sup> But we see no justification for this. Again, we are not here in a position where we must select just one alternative, and, if we are interested in obtaining as much positive utility as we can, it is only reasonable to accept all those sentences as hypotheses whose expected utility is both maximal and positive. To do less would be to pursue a miserly strategy in the search for truth.

However, even if we do accept all those sentences as hypotheses by induction whose expected utility is both positive and maximal, we shall often allow the truth to escape us when there is no need to do so. Our policy remains insufficiently aggressive in the pursuit of truth.

This argument is best illustrated with an example. Suppose a scientist is interested in the location of a particle on a grid when his evidence tells him that it is in one of the 10,000 squares in the center of the grid. Here the members of the defining partition

in the reduced competition set consist of sentences  $H_1, H_2$  and so forth to  $H_{10,000}$  affirming that the particle is in one of the center squares on the grid. Thus, for any  $H_i$ ,  $m(H_i, E) = .0001$ . Let us suppose that the evidence  $E$  expressed in terms of the marks left in square 432 indicates that the particle is in that square. Imagine that the conditional probability of  $H_{432}$  on the evidence  $E$  is .99989, and that, though the probability of each of the other hypotheses on  $E$  is greater than 0, the most probable among them is  $H_{78}$  which has a probability on the evidence of .0001001. In this case, since the expected utility of  $H_{78}$  is positive, it would not be reasonable by our rule to accept  $H_{432}$ . For, given the additivity of expected utility the expected utility of the disjunction,  $H_{78}$  or  $H_{432}$ , is greater than the expected utility of  $H_{432}$ . Hence, we would be prohibited by our equation for expected utility and the rule to accept what is both positive and maximal from accepting  $H_{432}$  as a hypothesis by induction from the evidence.

This is unsatisfactory from the standpoint of pursuing the goals of seeking truth and avoiding error. Sentence  $H_{432}$  is both maximally informative and exposed to a very small risk of error, smaller than any other maximally informative sentence. Indeed, the sentence has an extraordinarily greater chance of being true than any other sentence of the same content. Thus, no matter how high the probability of some maximally informative sentence of the reduced competition set may be, we are kept from accepting that sentence as a hypothesis if there is any other equally informative sentence whose expected utility is positive no matter how improbable on the evidence that sentence may be and no matter how miniscule the positive expected utility of it might be. For, the additivity of expected utility resulting from the equalities laid down above guarantees that the disjunction of any two maximally informative members of the partition each of which has some positive expected utility, however miniscule, will have a greater expected utility than any single one of the members. That is what produces the untoward consequence articulated above.

*A Revised Utility Assignment.* To obtain a more satisfactory account of the acceptance of sentences as hypotheses by induction from the evidence, we must redefine our utilities to bring them more closely in line with our objective of accepting sentences as hypotheses which, if true, are highly

<sup>7</sup> Isaac Levi, *Gambling*, *op. cit.*, pp. 84-85, and Risto Hilpinen, *op. cit.*, p. 106.

informative. By equating the utility of accepting a sentence as a hypothesis when it is true with the content of the hypothesis, we fail to adequately express our preference for accepting highly informative sentences as hypotheses. We can correct this oversight and improve our account by indicating that we value accepting a sentence as a hypothesis having the content of a member of the partition twice as much as one having the content of a disjunction of two such members, three times as much as one having the content of a disjunction of three such members, and so forth.

If  $H$  is a sentence in the reduced competition set, and  $h$  is the number of members of the partition disjoined in  $H$ , or equal to one if  $H$  is a member of the partition, we obtain the desired result by redefining the utilities as follows:

$$u(TH, E) = \text{cont}(H, E)/h$$

$$u(FH, E) = -\text{cont}(-H, E)/h.$$

Given these equalities, we obtain the equation:

$$e(IH, E) = \frac{p(H, E) - m(H, E)}{h}$$

which insures that any member of the partition that has an expected utility greater than the expected utility of any other member, which will result whenever such a member is more probable on the evidence than any other member, will also have a greater expected utility than any other sentence. If no such member has a greater expected utility than all the other members, then any disjunction of two members having a greater expected utility than all the other members will have a greater expected utility than any other sentence. Thus, in the example sketched above, the sentence  $H_{432}$  will have a greater expected utility than any other sentence. Expected utility as thus specified is not additive, of course, though the function in the numerator of the equation is.

We can, by algebraic manipulation, convert the present equation into one of greater philosophical interest. Notice that we could preserve the same ranking of expected utility of sentences in the reduced competition set if we multiplied the numerator of the right hand fraction by some constant, for example, by the number  $n$ , which is the number of members of the partition in the reduced competition set. We then obtain the following equation:

$$\begin{aligned} e(IH, E) &= \frac{(p(H, E) - m(H, E))n}{h} \\ &= \frac{p(H, E) - 2/m(H, E)}{h/n} \end{aligned}$$

Given the equation

$$m(H, E) = h/n$$

we obtain the following results:

$$e(IH, E) = \frac{p(H, E) - m(H, E)}{m(H, E)} = \frac{p(H, E)}{m(H, E)} - 1.$$

The ranking of the sentences in terms of expected utility will be determined entirely by

$$p(H, E)/m(H, E)$$

in the equation, and, given the equality of the antecedent probabilities of members of the partition in the reduced competition set, we may obtain the same ranking from the function

$$p(H, E)/p(H)$$

by an argument given earlier. Then we may, when the equality of antecedent probabilities in question holds, substitute the following equation:

$$e(IH, E) = \frac{p(H, E)}{p(H)} - 1.$$

The righthand side of this equation is a function which is a quotient measure of the relevance of  $E$  to  $H$ ; the value is positive if  $E$  raises the probability of  $H$ , negative if  $E$  lowers the probability of  $H$ .<sup>8</sup>

This measure of relevance yields the same values as the function

$$\frac{p(E, H)}{p(E)} - 1$$

which is a measure of the relevance of  $H$  to  $E$ ; again the value is positive if  $H$  raises the probability of  $E$ , negative if  $H$  lowers the probability of  $E$ . Hence, the measure of expected utility may be considered a measure of the power of  $E$  to support or confirm  $H$ , since it is a measure of the relevance of  $E$  to  $H$ , and it may also be considered a measure of the power of  $H$  to explain or anticipate  $E$ , since it is a measure of the relevance of  $H$  to  $E$ . Accepting a sentence as a hypothesis by induction from the evidence which has an expected utility that is both positive and maximal thus amounts to accepting one which the evidence has the greatest power to confirm and which has the greatest power to explain the evidence.

Unfortunately, given the current utility assign-

<sup>8</sup> Cf. Henry Finch, "Confirming Power of Observation Metricized for Decisions Among Hypotheses," *Philosophy of Science*, vol. 27 (1960), pp. 293-307, 391-404.

ment, it follows that inconsistent sentences may have an expected utility that is maximal and positive. Indeed, any member of the partition whose acceptance as a hypothesis has an expected utility that is as high as any other member of the partition will have an expected utility that is maximal. For the expected utility of accepting a disjunction of  $n$  such members will be equal to the sum of the expected utilities of those  $n$  members divided by  $n$ . Of course, two members of the partition whose acceptance as hypotheses is positive and maximal cannot both be accepted without falling into contradiction. Hence, in contrast to this the preceding utility assignment where it was reasonable to accept all those sentences as hypotheses whose acceptance was positive and maximal, the present assignment requires a rule for ties to escape from the certain error of inconsistency. The needed rule is easy enough to formulate. We need a method for breaking ties only when two or more members of the partition in the reduced competition set have a maximal expected utility. And, the rule should be to accept the disjunction of such members in numerical order. This disjunction will also have an expected utility that is both positive and maximal.

The utilities we have defined and the equation for expected utility resulting from them form the basis for a justification of the rule of induction (IR) and the rule of acceptance (RA) laid down earlier. If we follow the policy of accepting that sentence of the reduced competition set whose expected utility is both maximal and positive, or, in case of ties, the appropriate disjunction, and, because we are seeking truth, the deductive consequences of that sentence in the competition set, we obtain the results obtained from (RA). If we accept all those sentences in the reduced competition set as hypotheses by induction from the evidence whose expected utility is greater than other sentences of the same content, then we shall accept as hypotheses exactly those sentences accepted by (IR). This then completes our justification for those rules in terms of the theory of expected utility.

Since the decision theoretic rule of acceptance yields the same results as (IR) and (RA), we can conclude that any probability functor that yields the same comparative conditional probabilities will be equally satisfactory as a guide for the acceptance of hypotheses. All that matters is comparative probabilities, because a sentence will have an expected utility that is greater than other sentences of equal content if and only if it is more probable on the evidence than other such sentences. For the

purpose of accepting sentences as hypotheses by induction from the evidence as well as for the purpose of accepting sentences as evidence, all that matters is the comparative probabilities. Looked at from the standpoint of such acceptance, the choice of a quantitative probability assignment is an arbitrary choice from among an infinity of functors that yield the same comparative probabilities. Indeed, the notion of a single correct quantitative probability function is, for our purposes, no more than a useful fiction.

*Summary and Conclusion.* We have argued that the acceptance of sentences as evidence and as hypothesis by induction from the evidence is based on comparative subjective probabilities and a relation of competition among sentences. The relation of competition reflects our interest in seeking truth and avoiding error. We judge that the goal of avoiding error is more important in the selection of sentences as evidence, and, consequently, the competition for such acceptance is made especially rigorous. By reference to some touchstone hypotheses we first obtain a class of relevant sentences constituting a competition set, and then require that for a sentence within that set to be accepted as evidence the sentence must be more probable than any sentence with which it could possibly conflict. Once we have accepted certain sentences as evidence the competition set must be suitably reduced. The remaining sentences then constitute a competition set for the acceptance of sentences as hypotheses from the evidence. In accepting sentences as hypotheses the more important goal is that of seeking truth, though with some regard for avoiding error. Hence, we only require that a sentence in the reduced competition set accepted as a hypothesis by induction from the evidence be more probable on the evidence than others equal in content, that is, equal in the amount of truth about the world they contain if they in fact are true. Thus, while we require of sentences selected as evidence that they compete favorably, that is, be antecedently more probable, than any sentence of the competition set with which they could possibly conflict, we only require of sentences selected as hypotheses by induction from the evidence that they compete favorably, that is, be conditionally more probable on the evidence, than other sentences of the reduced competition set equal to them in content.

In conclusion, let us reiterate the salient features of these analyses. The acceptance of sentences as evidence or as hypotheses is based on comparative subjective probabilities. These are comparative

estimates by a person of the chances various sentences have of being true. The estimates may be formed and restrained by many factors. Intersubjective agreement, experimental test, conceptual innovation all play a role. But the bottom rock on which all else rests is the judgment of the individual investigator. The theory we have articulated places the ultimate burden of acceptance or rejection of sentences as evidence or hypothesis on the epistemological workhorse of individual judgment. Whatever guidance the individual receives from

the testimony of others and from experience, he must ultimately make up his own mind about whether, in the light of testimony, experience, and all else besides, this or that sentence has a better chance of being true. As he seeks truth and eschews error, what he accepts as evidence and hypothesis will depend entirely upon how he has resolved this basic epistemological question.<sup>9</sup>

*University of Rochester*

*Received December, 1972*

## POSTSCRIPT

(A Correction Added in Proof)

In the original version of "Truth, Evidence, and Inference," I offer an equivalence proof that requires elucidation. Suppose we have a set of sentences that constitutes a partition  $P$ . We then reduce that partition by deleting those members inconsistent with a statement of evidence  $E$  and obtain  $P_E$ . We formulate a competition set  $C_E$  of members of  $P_E$  and of disjunctions of those members in alphabetical order. For any sentence  $H$  in  $C_E$ , let  $h$  equal the number of members of  $P_E$  disjoined in  $H$  if  $H$  is disjunction, or one if  $H$  is a member of  $P_E$ , and let  $n$  equal the number of members in  $P_E$ . Letting  $p$  be a probability functor and  $e$  a functor for expected utility, I argued that

$$e(IH, E) = \frac{p(H, E)}{h/n} - 1$$

and that the same ranking of sentences in terms of expected utility results from the ratio  $p(H, E)/(h/n)$ . Earlier I had noted that if all members of  $P_E$  have equal antecedent probability, that is, for any such member  $M$ ,  $p(M) = s$ , where  $s$  is a constant, then, for some constant  $k$ ,  $p(H) = kh/n$ . Since the ranking of expected utility is not changed by multiplying  $p(H, E)/(h/n)$  by a positive constant, for example,

$1/k$ , it follows that the same ranking would be obtained from  $p(H, E)/(kh/n)$  and hence from  $p(H, E)/p(H)$ . This proof is correct. However, I had mentioned earlier, in connection with an equation for expected utility I rejected, namely,  $e(IH, E) = p(H, E) - h/n$ , that the same ranking may result if we replace  $h/n$  with  $kh/n$  and hence with  $p(H)$  in the equation. A different ranking may also result from this replacement, for example, if  $k$  is some very small number. In my proof that  $p(H, E)/(h/n)$  yields the same ranking as  $p(H, E)/p(H)$  on the condition that members of  $P_E$  have the same antecedent probability, I refer back to the discussion of  $p(H, E) - h/n$ , and it might appear that I was assuming that  $p(H, E) - h/n$  yields the same ranking as  $p(H, E) - p(H)$  under the condition mentioned. I do not make the latter assumption, and, indeed, it is false. The proof does not depend on that assumption, however, but on the weaker and correct assumption that for any sentence in  $C_E$ ,  $p(H) = kh/n$  for some positive constant  $k$ , on the condition that every member  $M$  of  $P_E$  is such that  $p(M) = s$ . For any such  $H$ ,  $p(H) = hs$  by the additivity of probability, and, therefore,  $k = sn$ .

<sup>9</sup> Research for this paper was supported by the National Science Foundation.

## II. GOD AND MYSTERY

J. KELLENBERGER

### I

PERHAPS central to every religion is a sense of mystery before the world. In the religions of emerging societies there is an awe before certain phenomena, seen as theophanies, or signs of gods; and in monotheistic religions we find a similar sense. The experience of mystery certainly is a part of the genesis of Judaeo-Christian religion. The Psalmist expresses this experience over and over. Thus he says in the 19th Psalm: "The heavens are telling the glory of God; and the firmament proclaims his handiwork," and in the 90th: "Lord, thou hast been our dwelling place in all generations . . . from everlasting to everlasting thou art God." Many religious persons taste this experience; although perhaps few drink fully from its cup.

In the history of theology there is a pendulum swing between doctrine and experience. At one extreme there are Aquinas, with his elaboration of doctrinal distinctions and apologetics, and Paley with his argument for the existence of God. At the other extreme there are the mystics and Schleiermacher with his stress upon an experience of dependence as the heart and capstone of religion. Within the Judaeo-Christian tradition each extreme is associated with a tendency. When the pendulum is at the doctrinal apogee the tendency is toward a system that fails to touch men emotionally or religiously. When the pendulum is at the experiential apogee the tendency is toward the elevation of experience to the exclusion of all belief, even belief in God. At present the pendulum seems to be swinging in the experiential direction.<sup>1</sup>

Schleiermacher spoke of a feeling of dependence; others speak of an experience of mystery. Perhaps what Schleiermacher identified as a feeling of dependence is the experience of mystery with an emotional overlay, as it were. In any case the generic or more central experience of religion is that identified by Rudolf Otto as the numinous experi-

ence, the religious feeling of profound awe and mystery.<sup>2</sup> Some have thought that religion is essentially this experience of mystery, and the competing beliefs of different religions stand opposed to this experience and hence to what is most truly religious.

In what follows we shall be concerned with two major questions about the nature of the religious life that it seems to me this thought and its present ascendancy make very pressing: How does the experience of mystery relate to religious belief, particularly belief in God? And how does it relate to religious commitment? Our concern is not with the variety of effects the experience of mystery historically has had on belief and commitment; nor is it with what a man who has had this experience might idiosyncratically bring himself to believe or not believe. Our concern is with the significance of this experience as such for the religious life of belief and commitment as such. In this sense our concern is not historical or psychological, but philosophical.

### II

It is one thing to say that the experience of mystery is at the heart of religion. It is another to say it is the essence of religion and belief is peripheral. And it is yet a third thing to say that it is the essence of religion and completely foreign to belief, which conceptualizes God. This it seems to many is what Otto thought and, following Otto, W. T. Stace elaborates this idea.<sup>3</sup> Stace feels it is no accident and essential to Christianity that at its center is the mystery of the Trinity. More than this, Stace feels that the mystery of the Trinity, like all religious mysteries, must embody the self-contradictory. If the paradox that God is One in Three were understandable, then, he believes, the mystery of the Trinity would be destroyed and with it Christianity. But if it is self-contradictory, it can

<sup>1</sup> On this see David Cairns, *God Up There? A Study in Divine Transcendence* (Philadelphia, 1967), p. 21.

<sup>2</sup> In *The Idea of the Holy*, trans. John W. Harvey (2nd edition: New York and London, 1950).

<sup>3</sup> He does so in *Time and Eternity* (Princeton, 1952). Cf. Otto, *The Idea of the Holy*, *op. cit.*, p. 2 and Thomas McPherson, "Religion as the Inexpressible," *New Essays in Philosophical Theology*, ed. by Antony Flew and Alasdair MacIntyre (New York and London, 1955). Also cf. Harvey's preface to the 2nd edition of *The Idea of the Holy*, *op. cit.*



never be rationally understood. Stace affirms that *all* propositions about God, taken as such, are self-contradictory or lead to self-contradiction and so are false. Propositions like "God is Trinitarian" are internally self-contradictory, while other propositions lead to a self-contradiction in conjunction with a second proposition ("God is a Person" in conjunction with "God is Infinite"). From this Stace infers that all literally said of God is false.<sup>4</sup> Now perhaps Stace ought not to infer this, since if a proposition is self-contradictory its denial is true, and if two propositions contradict each other, the denial of their conjunction is true. But maybe Stace thinks that from the very fact that a concept (that is, a predicate) is applied to God a self-contradiction somehow follows or ensues, since, as he believes, God is "beyond all conception."<sup>5</sup> In any case, Stace is quite clear in *Time and Eternity* that all literally said of God is false, and hence there can be no true beliefs about God.

Appreciating that if this is so religious language must be used exclusively for something other than description (or at least true description), Stace says that religious language is exclusively evocative—its purpose being to evoke the experience of mystery.<sup>6</sup> He would, I think, take the formally contradictory utterances of the mystics to be paradigmatic of religious evocation. Examples might be:

"God is near and not near."

"God was born in Christ, yet He was not born."

"God is temporal but eternal."

"Thou shalt love God as He is,

a Non-God. . ."

This last is drawn from Eckhart and supplied by Stace.<sup>7</sup> Surely the mystics used such expressions, and doubtless Stace is correct that they were used to evoke experiences in others. Further, such paradoxical expressions have increased evocative power by merit of their paradox. But is Stace right that such expressions are all false? Or, more exactly, is

he right in his implication that religion would be destroyed if such expressions, or some other descriptions of God, turned out to be true? I think not, and I think that the very mystics Stace cites would differ with him on this implication and would take their paradoxical utterances to be true.

To see how mystics could take their paradoxical and self-contradictory utterances to be true we need only remind ourselves of the distinction between *contradictions in meaning* (containing contradictory or contrary propositions) and *contradictions in form*. It is the former that are real contradictions, the latter being contradictions only in appearance. In the proper context this would be an example of a contradiction in form: "That man is not a man; he is an animal." Or, again: "Our life together has been full; yet it has been empty." Both of these statements look like contradictions, but understood, could well be true. My suggestion is that, similarly, the paradoxical utterances of mystics are formal contradictions in this sense and, if understood, could well be seen to be true—or at any rate to be not necessarily false. In this way the mystics' paradoxes could both evoke and describe.

Indeed, sometimes mystics themselves point to the way to understand their paradoxical, contradictory utterances. Sankara, Stace tells us, tried to resolve how Brahman could both have qualities and not have qualities, both be and not be, by distinguishing between a "higher" and a "lower" Brahman.<sup>8</sup> Stace objects to Sankara's distinction because, he says, it denies that God is One, which holds alike for Hinduism and Christianity. I do not think that Stace is right here. I do not think that distinguishing between a "higher" and "lower" Brahman denies that Brahman is one, for such a distinction may be understood as being between two *aspects* of the one Brahman. In a similar way, regarding the Christian Trinity, God the Father may be regarded as one aspect of the One God, and God the Son another, which allows that there are two Persons, but one Being.<sup>9</sup> How-

<sup>4</sup> Stace, *Time and Eternity*, *op. cit.*, pp. 8 and 60-61.

<sup>5</sup> Stace, *ibid.*, p. 61. In *Mysticism and Philosophy*, however, Stace insists that concepts *do* apply—at least to the mystic experience. That is, he says, the mystic can correctly and literally say of his experience that "it is *x*" and "it is not *x*," *Mysticism and Philosophy* (Philadelphia and New York, 1960), p. 304.

Thus for the Stace of *Time and Eternity* all beliefs about God are false, while for the Stace of *Mysticism and Philosophy* all beliefs about God, or the experience of God, are "true," including contradictory beliefs. But this clearly is a novel and necessarily nonliteral sense of "true." Our concern will be with Stace's view in *Time and Eternity*.

<sup>6</sup> Stace, *Time and Eternity*, *op. cit.*, p. 86. In this he echoes Otto.

<sup>7</sup> *Ibid.*, p. 10.

<sup>8</sup> *Ibid.*, p. 14.

<sup>9</sup> On how aspect talk relates to the Incarnation see Robert Herbert, "The God-Man," *Religious Studies*, vol. 6 (1970), pp. 157-174.

ever, there is something more important to note about Stace's point that what Sankara says denies that God is One. To make this point—which is not simply that what Sankara says contradicts "God is One," but rather that it wrongly denies that God is One—Stace must assume that *it is true* that God is One. In short, to make this point Stace must run counter to his thesis that all descriptions of God are false.

One mark of the fact that mystics did not regard their paradoxical, contradictory utterances as real contradictions is that they regarded them as true, as important religious truths—just as Stace regards it as a truth that God is One. When one reads Meister Eckhart's sermons one senses that Eckhart was trying to move his hearers, to evoke something in them, but also that he was trying to communicate important truths. In one sermon he cries, "Oh—if only the soul's eyes were opened so that it might see the truth!"<sup>10</sup> In that same sermon he uses contradictory expressions, but expands them to make clear his meaning. He says: "There is an agent of the soul—no, not an agent but something more, a Being—yet not a Being. . . ." And then he goes on to explain that it is not a Being in that it is utterly uncreaturelike, but it is a Being in that it is God Himself dwelling in us.<sup>11</sup>

Of course sometimes when a paraphrase or an expansion is added to the mystic's formally contradictory expression, while the truth or meaning he wants to communicate is clarified, its verve and evocative power are lost. Mystics, or mystics like Eckhart anyway, wanted primarily to bring others to see what they had seen. They did not want *simply* to state the truth they had before them. They did not because they realized that they had to do more than say what they had seen to get others to see it; they had to evoke a change in them and bring them to see it for themselves, they had to "open the soul's eye." And for this task evocation, not statement, must do service. But it does not follow at all that they therefore expressed nothing true in their evocative language. Consequently, even if no expansion of a mystic's paradoxical utterance is found in his writings, this does not mean that what he said is false because self-contradictory, or that he thinks it is. Moreover,

even if all the mystics' paradoxical and formally contradictory utterances were indeed contradictory, and so false, that would not mean that there were *no* true or false descriptions of God, or that any mystics thought this. Discounting the paradoxical utterances of the mystics—which I have suggested Eckhart at least considered to be truths properly understood—there would remain the things said of God that are not formally contradictory, such as, that God is merciful and that He created the world. Let us recall that Christian mystics, as Christians, believed the central articles of Christianity to be true. Certainly St. Teresa of Avila did, and Meister Eckhart did too. Eckhart apparently was appalled at the thought of heresy and, albeit under inquisitorial pressure, proclaimed himself ready to retract any erroneous statements he may have made in his Sermons and writings.<sup>12</sup> And this is to say, first, that he thought there could be falsity or truth in our statements about God and, second, that he thought his own statements mostly or wholly true.

It seems, then, that contradictory mystical utterances are contradictory in form only. At least this is the case for Eckhart in the instance cited. Hence, at least sometimes mystics regard their contradictory utterances as true. Further, it does not seem that all other propositions about God must be regarded as false. At any rate, mystics, among others, take many of them to be true. Consequently, religion seems quite compatible with there being truth in some of the things said of God.

### III

What then of mystery? If there are religious truths, does that banish mystery? That is, if the religious person allows that there are religious truths, need that banish his sense of mystery before the world? I think not, as some attention to religious contexts shows us. I do not mean liturgical contexts, but those situations in which a sense of religious awe traditionally plays a part. More than one man has stood before the world and asked that logically odd but penetrating question "Why should there be anything?" Not all admit that this question makes good sense, of course.<sup>13</sup> But still

<sup>10</sup> R. B. Blakney, *Meister Eckhart* (New York and London, 1941), p. 134.

<sup>11</sup> *Ibid.*, p. 135.

<sup>12</sup> *Ibid.*, p. xxiv.

<sup>13</sup> J. J. C. Smart and Bertrand Russell do not, for instance. See J. J. C. Smart, "The Existence of God," *New Essays in Philosophical Theology*, *op. cit.*, p. 46; and the Russell-Copleston debate in Russell's *Why I Am Not a Christian* (London, 1957), reprinted in *The Existence of God*, ed. by John Hick (New York, 1964), where the pertinent pages are pp. 167–178.

men ask it. And many, if they answer it, do so in terms of the Numinous. They find themselves in the presence of God and to be part of God's creation, as they come to say. The question here is not whether they do in fact find themselves in the presence of God, but whether their believing this to be true need destroy their sense of mystery. I think clearly not. Indeed to have a sense of being in God's presence is to have a sense of awe. But to have a sense of being in God's presence is to believe that one is. It is to have what may be an overwhelming conviction that one is in the presence of God, which is to say it is to believe that it is surely true one is in God's presence. In this instance the sense of mystery before the world is, or comes to be, inextricably bound up with a belief.

In other religious situations men express or exhibit other beliefs that are compatible with a sense of God's mystery. In prayer God may be thanked for His constant mercy and goodness. A man who thanks God for His constant mercy clearly believes that God has been merciful to him, and that God is merciful. But this belief—that God is merciful—says nothing about the *way* of God's mercy. God's ways, we are told, are beyond our understanding. Here then is the element of mystery. It is regarding the mode of God's mercy, goodness and His other attributes. But *that* God is merciful and *that* He is good (and more) are taken to be truths by the man who thanks God for His mercy and goodness. This is a logical point. And it is one with a religious implication. The man who publicly utters thanks for what others see as God's goodness, while he himself cannot quite believe that it is goodness, comes as close to hypocrisy as a man in the analogous secular situation comes to insincerity. Yet, while the man who thanks God for His goodness must believe God to have shown him or another His goodness, he can still confess his lack of understanding and stand in awe before the mystery of God's ways. Job, it will be recalled, was rebuked out of the whirlwind for assuming he understood God's way, but his faith in God's goodness was approved.

Upon reflection, it becomes clear that even Otto made affirmations about the Numinous. Both Otto and Stace accept it that God is "beyond conception" and that "nothing can be said of God." But to say this, after all, is itself to say something of God. So even for Otto and Stace God can be "con-

ceptualized" and there can be true beliefs about Him. At this juncture the impulse of some is to fall silent. But this is no solution. Ceasing to say what can be said of God hardly entails that nothing can be said of Him. Of course there is a problem here in that it seems it will be true that nothing can be said of God only if it is false. This oddity is an intimation of the tangle that can accompany saying that God is "beyond our concepts."<sup>14</sup> There are ways to cut through this tangle, however. Perhaps Otto and Stace would say that while it is a truth that nothing can be said of God, it is a meta-truth, not about God, but things said about God or, rather, thought to be said about God. Yet this will not do, for Stace, as we have seen, also believes that it is a truth that God is One; and this is a truth about God. Otto has a longer list of truths about God: for Otto it is a truth that it is the Numinous or God that evokes the sense of mystery he discusses; and that it is the same Numinous that evokes the sense of majesty he discusses; and it is a truth that the Numinous is "wholly other," that is, "beyond the sphere of the usual, the intelligible, and the familiar, which therefore falls quite outside the limits of the 'canny', and is contrasted with it, filling the mind with blank wonder and astonishment."<sup>15</sup> This last truth says what God is *not* to some extent, but still is a truth about God.

My point here is not that Otto and Stace are wrong. It is that accepted truths about God need not destroy the sense of mystery in religion. Even Otto and Stace accepted some such truths. Were they wrong, then, in saying that God is "beyond conception"? I think not. Although what this means cannot be that virtually nothing true of God can be said. What it could mean is pointed to, I think, in a line quoted by Otto:

"A God comprehended is no God."<sup>16</sup>

God would cease to be God if He were comprehended, but God would not be comprehended if even a number of truths were known about Him. We do not fathom God if we appreciate that He is One, or "wholly other," or merciful, or good. For in understanding this much we do not understand His nature, His ways, the modes of His attributes of goodness and mercy. God could be "beyond conception," then, in the sense that His ways are beyond our conceiving, and just in this way He would remain mysterious.

<sup>14</sup> See William P. Alston, "Ineffability," *The Philosophical Review*, vol. 65 (1956), pp. 506-522.

<sup>15</sup> Otto, *The Idea of the Holy*, *op. cit.*, p. 26.

<sup>16</sup> Otto uses this quotation from Tersteegen as an epigraph to ch. 5 of *The Idea of the Holy*.

## IV

Let me raise a slightly different question: Does one who has not a clear conception of God necessarily lack in religious faith, feeling, and dedication? Kierkegaard in his *Concluding Unscientific Postscript* helps us answer this question while dealing with a related concern. He draws a contrast between two men. One "lives in the midst of Christianity [and] goes up to the house of God, the house of the true God, with the true conception of God in his knowledge, and prays, but prays in a false spirit." The other "lives in an idolatrous community [but] prays with the entire passion of the infinite, although his eyes rest upon the image of an idol." Kierkegaard wants us to see that the religiously important thing is praying with a true spirit, that is, having faith and dedication; it is not having the true conception of God. And, he wants us to see, one can pray in a true spirit, and have faith, even when he has the *wrong* conception of God. (Analogously, one man can have deep and committed faith in another human being even when he thinks of him as he is not—has the "wrong conception" of him.)

Now note that the idol worshipper had not only a wrong conception of God, but a clear one—God is the stone on which his eyes rest. Surely if a wrong clear conception of God need not destroy a man's religion, a right clear conception of God need not either. The mystics warn us about seeking to intellectualize and to rationalize God. But conceiving of God *per se* is not incompatible with religious faith and dedication.

Nor need having a certain conception of God lessen one's sense of mystery. A clear notion of God can be drawn from the Bible—all-powerful creator of the universe who is the embodiment of perfect love. Take the conception of God as creator. The sense of mystery before the world need not be lessened by conceiving of God as creator and seeing the world as His creation. Indeed seeing the world as the creation of God is one way of seeing it as a place of awe. The Psalmist expresses his awe before the world *by* saying that the heavens show forth the glory of God and the firmament His handiwork.

But there is a danger here that Otto, Stace, and others help us see. If a man dwells upon a conception of God, and its logical features, in the absence of religious feeling, then his religion becomes an intellectual exercise. There is a sense in which

religion is grounded in experience, and Otto and Stace remind us of this sense.

## V

However, there is another danger in what Otto and Stace say. It is that religion will become identified with having the numinous experience, and God will be equated with the experience itself. This in fact is just what Stace does.<sup>17</sup> Thus for Stace the experience of mystery is not *before* God or *of* God, but *is* God. "What, though, is wrong with this?" one might ask. "Does it not make religion and God nicely empirical, and does it not open up religion to all men?" If it makes religion empirical, it does so only by remaking God and, as well, the nature of the religious life which draws its sustenance from God. Such a God could not be thanked for His mercy, for God as an experience would not be the sort of entity that either does or does not show mercy. God would not be the creator of men, nor give grace to men, nor hear prayers; there would be no turning to God—none of these predicates have any sensible application to an experience.

But would not religion, as the numinous experience, be accessible to all men? This question leads to another question: Is the numinous experience possible *in vacuo*? That is, could a man take the experience by storm, as it were, regardless or in spite of all save the sensations he came to have during the experience? Could a man, regardless of all else, induce in himself the numinous experience with lysergic acid, for instance?

Huston Smith in his article "Do Drugs Have Religious Import?" says evidence indicates "that chemicals *can* aid the religious life, but" he adds, "only where set within a context of faith (meaning by this conviction that what they disclose is true) and discipline (meaning diligent exercise of the will in the attempt to work out the implications of the disclosures for the living of life in the everyday, common-sense world)."<sup>18</sup>

Ecstatic experiences, as such, are just that: ecstatic experiences, neither more nor less (whether drug induced or not). The experience of bliss or ecstasy, or any other psychological experience, can comprise the whole of only a truncated religion. They cannot, as such, be equated with the numinous experience. They cannot because if we can say nothing of an experience beyond that it is an

<sup>17</sup> Stace, *Time and Eternity*, *op. cit.*, p. 76-77.

<sup>18</sup> Huston Smith, "Do Drugs Have Religious Import?" *The Journal of Philosophy*, vol. 61 (1964), pp. 529-530.

experience, then it can play no role in a *context of faith and discipline*. We do not have faith in an experience, but in that which the experience is of or taken to be of. We do not try to live by an experience, but by that which is disclosed in the experience or taken to be disclosed. In this way the two elements Smith points out—faith and discipline—indicate that to be numinous a drug induced experience, or any, must relate to what is beyond the experience itself.

This finding accords with what Otto himself says. Otto speaks of "the object to which the numinous consciousness is directed." He calls this object the "wholly other," and for Otto it is this object *before* which we feel awe and mystery in the numinous experience.<sup>19</sup> Observe that religious awe is *before* something. All awe is before something, and so too for religious awe, where the object is the numinous object. This is not to say that if men have had the numinous experience (as men have), their experience necessitates the existence of the numinous object; it is to say that their experience is directed beyond itself to an object which is, in Otto's words, "felt as objective and outside the self."<sup>20</sup> This means that we cannot identify the religious experience of awe or mystery purely phenomenologically, that is, in terms of psychological and emotional traits that do not point to anything beyond themselves. If we include awe in a phenomenological description of an experience, then we must bear in mind that it points beyond itself in the way indicated, whether to the numinous object or to something lesser. Can a drug produce the numinous experience, then? The answer seems to be sometimes drugs do, sometimes they do not. For the participants in the Marsh Chapel experiment, it may well be that psilocybin did.<sup>21</sup> But in any case if a drug-produced experience does not contain awe before the "wholly other," but only ecstasy, changed sense perception, a feeling of well-being and so on, the produced experience is not the experience of the Numinous.

If it is not already apparent, this necessity of awe for the numinous experience, and the necessity of this awe being before the "wholly other," carry an implication for religious belief. Making it explicit, it is this: if a man feels religious awe he must

believe that, in some way, he is before the Numinous. That is, he must believe that in some way he has confronted the "wholly other"—something by its very nature distinct from himself and his experience, but which elicits from him his feeling of profound awe. And so it emerges that if the numinous experience has as an essential element awe before the "wholly other" (and it does), Otto's numinous experience *requires* the conviction of belief—the belief that the Numinous has been encountered. Thus Otto's experience not only is compatible with religious belief, but it demands at least one religious belief.

This necessity of belief for the experience of the Numinous is brought out in an episode in *The Teachings of Don Juan* by Carlos Castaneda.<sup>22</sup> At one point Don Juan and Castaneda are engaged in the ritualistic gathering of peyote buttons, and Don Juan warns Castaneda to approach the cacti reverentially, bearing in mind what they signify, namely, the presence of Mescalito. Castaneda has told us of his experiences with peyote, in which he has seen a being who appears either as a man with green, knobby skin or as a light. But Don Juan apparently suspects that Castaneda regards these experiences essentially as hallucinations. For Don Juan these experiences are of the Numinous, of the "wholly other" in the form of Mescalito. He sees the visual experiences as vouchsafed by Mescalito, even though they are induced by the peyote buttons. Castaneda, on the other hand, has an attitude toward them more appropriate toward an interesting and vivid sensation than toward a confrontation with the "wholly other." And this is to say that he has not in his mind confronted the Numinous, as Don Juan has. Castaneda will have the Numinous experience when he ceases to regard Mescalito as an hallucination, and feels before his vision the sense of awe and mystery one feels before the Numinous. But this awe will come only with the belief—the realization one with the experience would say—that there is something before which to feel the profound awe of the Numinous. Again, the two are one: to feel awe *before* the Numinous is to feel one *is* before the Numinous—that is, to believe with conviction that one is before the Numinous. So long, then, as this

<sup>19</sup> Otto quite rightly differentiates between awe and mystery. Sometimes, he says, a sense of one preponderates in the numinous experience, sometimes a sense of the other, *The Idea of the Holy*, *op. cit.*, p. 25. For the sake of simplicity I shall for the most part carry on the discussion in this section in terms of awe alone.

<sup>20</sup> Otto, *The Idea of the Holy*, p. 10.

<sup>21</sup> See Smith, "Do Drugs Have Religious Import?" *op. cit.*, p. 521.

<sup>22</sup> Carlos Castaneda, *The Teachings of Don Juan* (New York, 1969), pp. 84–103.

belief is lacking there cannot be the encounter and Castaneda's experience will remain at the level of an absorbing and frightening visual hallucination. Even if his experience were identical phenomenologically with Don Juan's (which it apparently approximates), it would fail to be of the Numinous without the sea change of his acceptance of Mescalito as the "wholly other." Through these reflections as well, then, we can discover in the nature of the numinous experience the necessity of belief in something beyond the experience itself.

## VI

So far I have tried to show how the sense of mystery before the Numinous is not only compatible with certain religious beliefs, and having a clear conception of God, but actually requires certain beliefs. Let us go a step further: not only religious belief, but religious knowledge is compatible with religious mystery. One reason the sense of religious mystery can arise is because God's *ways* of goodness, mercy, and more are beyond our conceiving; at the same time *that* He is good and *that* He is merciful are believed and presupposed as beliefs in, for instance, the religious activity of thankful prayer. This source of mystery—the inconceivability of God's ways—would remain just as it is if it were *known that* there is a God, *that* He is good and *that* He is merciful, and hence the mystery would remain. Anselm, who felt mystery and awe before the Lord, nevertheless thought that it had been given to him to prove, and so to know, that God exists.

However, there is another point of view here. It is a point of view both penetrating and blinkering. Roughly, what Otto and Stace suggest is that God's *nature* is mysterious, which allows that God can be believed to exist and even known to exist. The burden of this other point of view is that it is God's *existence* that is mysterious; it is mysterious in that whether there is a God can never be evident to men. And this is to say that men cannot know that there is a God. One might think that the basis for this view is that man's mind is limited and incapable of gaining such knowledge. But not so; the basis for the view has to do with the implications of such knowledge if it were gained. It is contended that if men knew there is a God, then they could not choose to recognize God; and this would mean that no personal relationship with God would be possible.<sup>23</sup>

It is to the credit of this view that it reminds us of the central place in Judaeo-Christian religion of a personal relationship with God. Philosophers' almost ubiquitous use of the word "theist" makes it seem as though a believer were some sort of theoretical thinker who is arguing for the "position" that God exists. In fact, the believer has belief *in* God, which involves a relationship of trust, and, if God is the perfect embodiment of love, a relationship of love between men and God. It is true that one who believes *in* God believes *that* God exists, and it is true that many religious thinkers have tried to prove the existence of God. But at the same time the typical concern of believers in the Biblical tradition has been to maintain their relationship to God of faith and trust and to deepen the relationship of their love for God; it has not been to advance the proposition that God exists.

When we come to the rest of the view, however, there is some confusion. For one thing, what is chosen? Is it to *recognize* God, to become *aware* of Him? Is it to *believe* in God? In fact, although the logic of these concepts is very complex, it seems that neither awareness of God nor belief in God must be *chosen* by the believer. Let me try to bring out why this is so by sketching in what seem to me to be salient features of awareness and belief. I shall then go on to another element that does require choice, but does not require that God's existence be mysterious in the sense of not evident.

To begin with, awareness as such does not seem to be a matter for choice. That is, neither choice nor coercion seems to relate to it directly. When I look up and become aware of the wind moving the uppermost branches of the trees, I am neither forced into that awareness nor choose it. Force, it seems, indirectly relates to awareness only in the following tenuous way. A man might be put into a situation or made to assume a position where he could not avoid becoming aware of something. We might say here that he was forced to become aware of that thing. Strictly, though, this means only that he could not avoid the recognition he made. And even this attenuated sense of "force" will apply only if he has discovered what he did not want to discover. Further, if he had not been put into that situation, and had made the discovery on his own, he would not have chosen to make it—any more than I choose to discover the wind in the trees. If "recognize" is allowed to mean "concede," that is another matter. Making a concession or admission, if it is an admission one does not want to make,

<sup>23</sup> Cf. John Hick, *Faith and Knowledge* (2nd edition: Ithaca, N.Y., 1966), pp. 120ff.

is something one might be forced to do or choose to do. This point, I think, relates to the fact that *making an admission*, like *believing* and unlike *becoming aware that something is so*, is volitional, that is, is something one can try to do or try not to do.

Belief is rather different from awareness. One can in the full sense of "force" be forced to believe something, it seems. A person can be forced by undeniable evidence to believe something against his will, that is, when his inclination and effort are not to believe it. And a person can choose to believe, just as he can try to believe or try not to believe. However, while this is so, it does not follow that one either chooses or is forced to hold every belief one has.

If we consider the Biblical man of faith, Job, who had a personal relationship with God if any man ever had, I think that we can plainly see he did not choose his belief in God, nor, of course, his awareness of God. Job, at least in his own eyes, lived each day of his life in the presence of God. God was not hidden to Job but before him in every element of His creation. God visited him in the bounty he reaped from his fields, and He visited him in his suffering and affliction. This nearness of God Job accepted as the most significant, and yet the most apparent, fact of his life.

For Job God's presence was certain. His cry of faith was not "I believe my Redeemer liveth," but "I know my Redeemer liveth." Job, then, did not choose to believe in God, for a person does not choose to believe what is certain and proclaimed on every side. A person chooses to believe only what is not established as certain as he sees it. Nor was Job's awareness of God chosen since, if I am right, awareness is never chosen. At the same time Job was not forced to believe in God since he was not forced to believe what he did not want to believe. In what might approximate the language of the Job-like man of faith, it was given to Job to be aware of God and to believe in Him.

As long as Job is the paradigmatic man of faith, belief is not something that is chosen. But there are two other contrasting models of faith that do portray the man of faith as one who chooses to believe. These are "contemporary models and show an existentialist influence. The first I shall call the Absurd model and the second the Paradox model.

The man of faith according to the Absurd model is one who believes in God, but believes against all reason. Like Job he looks about the world, but

unlike Job he does not detect God's presence on every side. Instead he encounters a vast silence, or, worse, he finds God a myth and that the world speaks against His existence. Yet he believes. He believes, in Kierkegaard's phrase, "by virtue of the absurd." Job's cry of faith was: "I know my Redeemer liveth." The cry of faith here is: "It cannot be, but I believe." Here the man of faith chooses to believe, and does believe, in the face of no evidence and even negative evidence against there being a God.

With the Paradox model the man of faith does not have faith; yet he tries with every fiber to believe, and so affirms the faith he does not have. His cry of faith is: "I cannot believe. God have mercy upon me." Miguel de Unamuno's story "Saint Manuel Bueno, Martyr" has as its protagonist such a Paradoxical man of faith; at any rate the story and Don Manuel are open to this construction. Don Manuel is the priest in a small Spanish village, Valverde de Lucerna. He is well loved by those in the village and known as a saintly man, not only kind but inspired in his kindness. Yet, while he helps others to believe, he himself cannot believe. He falls silent when he reaches the last verse of the Creed; although this is unnoticed in the village. He trembles when he gives Communion, but this too is unnoticed. Because his life is selfless, his secret—that he cannot believe—goes undetected. Here the man of faith chooses to believe, and tries to believe, but cannot.

Some place or other Unamuno has said, "Those who believe they believe in God but without passion in their heart, without anguish of mind, without uncertainty, without doubt and even at times without despair, believe only in the idea of God, not in God himself."<sup>24</sup> The elements of this description are met by the Absurd and the Paradoxical men of faith; that is, they have passion, anguish, uncertainty, doubt, and even despair. But the Biblical man of faith would fail to be a believer in God according to Unamuno's test, for while he has passion and anguish, he has no doubt or uncertainty. Kierkegaard says of Abraham that if he had doubted, he would *not* have had faith, and of course this is right—as far as the Biblical model goes. (Or at any rate his faith would have been lessened if he had doubted.) For Job and Abraham their commitment of faith is tied to their certainty and what they deem to be their knowledge.

Also, of course, while Don Manuel, the Paradox-

<sup>24</sup> Quoted by J. R. Jones in his "Love as Perception of Meaning," *Religion and Understanding*, ed. by D. Z. Phillips (Oxford, 1967), p. 148.



ical man of faith, has passion, anguish, doubt, and despair, he fails to believe. He fails to believe unless, as Kitty in *Anna Karenina* wants to say of Levin, he believes without knowing it. This is a fairly curious suggestion, perhaps; but oddity is at home in paradox. In any case, Unamuno has the narrator of his story make precisely this suggestion regarding Don Manuel. This notion may seem to run counter to the implication, discussed earlier, that believing in God presupposes believing that there is a God. It does not, for whether or not one must be aware of his belief in God, if one believes in God, he still must believe that there is the God in whom he believes—again, whether or not he is aware of it. In the epilogue to "Saint Manuel" Unamuno, speaking for himself, offers a less paradoxical, more balanced construction than his narrator, namely, that Don Manuel lived by works in the absence of faith. In any case, setting aside the question of whether the Paradoxical man of faith fails to believe or believes unawares, I think we can see that both the Absurd and the Paradoxical men of faith choose to believe, while the Biblical man of faith does not.

Let me try to sum up the import of this for the role of mystery, that is, uncertainty, regarding God's existence. All of these men of faith—the Biblical, the Absurd, and the Paradoxical—have in different ways a committed faith, or at least a committed effort to attain faith. The Absurd and Paradoxical men of faith, seeing God's existence as less than certain, if not against all reason, choose to believe; while the Biblical man of faith, seeing God's existence as evident, does not choose to believe, but believes as a man believes in what he could not deny if he wanted to. Belief in God, then, can be a matter of choice and may be chosen by many men of faith, but it need not be, for it is not chosen by the Biblical man of faith, for whom God's existence is as evident as his own life. For the Absurd and Paradoxical men of faith the means of their affirming and entering into a relationship of faith with God is their choosing to believe. Yet the Biblical man of faith Job had a personal relationship of faith and trust to God if any man ever had. And so, finally, a chosen belief is not necessary for a personal relationship to God, and thus there may be a personal relationship to God even if God's existence is evident, as Job conceives it to be, and in this sense not mysterious.

However, if belief may be born in the man of faith without being chosen, there is another element, equally important religiously, that must be

chosen. This element is the way the believer leads his life. Like faith, it too is a form of commitment. In the Judaeo-Christian tradition this form of commitment is made by "walking in God's way," or by "renouncing the world" or by "the imitation of Christ." Commitment, when it takes this form, is volitional in that it requires the choice, as well as the effort, to live a certain way. As such it stands in contrast to belief. Does this form of commitment require mystery, that is, uncertainty, regarding God's existence, even if faith does not?

Both Job and Don Manuel live religiously committed lives. Job is a "blameless and upright" man who walks in God's way. And Don Manuel leads a "saintly" life devoted to the people of Valverde de Lucerna. In fact Unamuno takes some pains to present Don Manuel as a Christ figure—an unbelieving Christ figure whose actions and compassion nevertheless imitate Christ's own. I take it here that Don Manuel did *not* believe, as Unamuno himself suggests. So the story of Don Manuel points up one more paradox: to walk in God's way one need not believe in God. This is what humanists who renounce their belief in the Biblical God, but look to the Bible for their moral precepts, have said all along. The primary point, though, is not that Don Manuel *can* choose to walk in God's way while not believing, but that both Don Manuel and Job similarly *must choose* to walk in God's way if they do so.

So works, or turning toward God in an active life, is over and above, or alongside, faith. And it must be chosen and actively entered into, as opposed to faith, at least faith in Job's case, which is given to him as it were, not chosen. Now if we grant this—that works must be chosen—which I think is surely true and ought therefore to be insisted upon, have we found something that necessitates there being uncertainty regarding God's existence? We have not. It is true that being certain there is a God is not turning toward God in one's life. In the same way being certain there is another who loves one does not come to loving the other. But it does not prevent it either. Analogously, certainty that there is a God need not prevent a man from turning toward God in his life. And this means that mystery, in the sense of uncertainty regarding God's existence, is no more necessary for this form of commitment than it is for faith. Don Manuel turned toward God in his actions, believing there is no God. But Job said he knew and was certain there is a God; he did not for this reason fail to turn toward God in his actions.



## VII

I have tried to show how the religious experience of mystery by its very nature cannot be alone at the center of religion, how it overlaps two other religious foci, belief and commitment, which themselves overlap one another. The experience of mystery—the numinous experience—far from being antithetical to religious belief, presupposes the belief that there is the “wholly other” before whom one feels mystery and awe. The numinous experience, in just this way, interlocks with faith. And faith, or religious belief, overlaps with commitment, is itself a paramount form of commitment.

Many of the questions that we have touched upon relating to mystery, faith, and commitment have direct implications for the religious life. For this reason in particular they cry loudly for resolution. Can one live a full religious life suffused in a sense of mystery without some form of commitment? Ought the religious person to shun belief as foreign to what is most truly religious and to seek experience without belief? Ought he to seek understanding of his belief or does understanding leach religion of its substance? Does belief require a choosing to believe, which in turn requires a lack of certainty regarding whether there is a God?

The confusions that prevent us from finding the answers to these questions are conceptual in nature to the extent that they revolve around what it is to

believe and what it is to experience religious mystery. Perhaps for some they are more than conceptual in the way that human problems and their roots often are. But to the extent that these confusions arise from unacquaintance with the structure of belief and mystery philosophical reflections can help toward their dissolution.

Earlier I observed, as others have, that the contemporary tendency in theological and religious thinking is toward a stress on experience to the exclusion of belief. This is one point. Mircea Eliade and Huston Smith make another point. Eliade in *The Sacred and the Profane* says that for contemporary nonreligious man there is no “properly religious experience,”<sup>25</sup> and Smith observes that “the phenomenon of religious awe... seems to be declining sharply.”<sup>26</sup> Both their point and the other are right, I think. Certainly they are compatible. There is, I suggest, even a connection between the two points. If I am right, the numinous experience requires belief in the “wholly other”; but if it is wrongly conceived of as *not* involving this belief and its commitment, then it could seem to deserve theological stress, not only as the touchstone of religion, but as the most accessible element of religion in an unbelieving age.

California State University, Northridge

Received January 9, 1973

<sup>25</sup> Mircea Eliade, *The Sacred and the Profane*, trans. Willard R. Trask (New York, 1957), p. 212.

<sup>26</sup> Smith, “Do Drugs Have Religious Import?”, *op. cit.*, p. 530.

### III. BERGMANN ON PERCEIVING, SENSING, AND APPEARING

DAN CRAWFORD

IN this study I am going to present and discuss some of the central themes of Gustav Bergmann's theory of perception. I shall be concerned, however, only with "later Bergmann," that is, with the perceptual theory worked out in a series of essays in which Bergmann shifts from phenomenism to a form of *intentional realism*. This label ("intentional realism") indicates the two dominant themes in Bergmann's later thought about perception: perceivings are analyzed as mental acts (thoughts) which are intentionally related to real and mind-independent objects and states of affairs.

In a timely essay, "Intentionality" (1955),<sup>1</sup> Bergmann presented an impressive defense of mental acts, and although the framework of that essay was still phenomenistic, the structural analysis of mental acts and their intentionality presented there has not been significantly altered by Bergmann in later writings. In two subsequent essays, "Acts" (1960) and "Realistic Postscript" (1963) his realistic turn was worked out in detail in the context of giving a satisfactory account of ordinary perceptual experience. These three essays will comprise the core texts of the study that follows.<sup>2</sup>

Bergmann's perceptual theory has not received its fair share of attention from the majority of contemporary analytic philosophers of perception. I can think of two reasons why this is so. First, there is the problem of the *approachability* of Bergmann's theory. His two most important essays on perception, "Acts" and "Realistic Postscript," are rather obscurely situated in a collection of essays which contains a formidable line-up of ontological and historical topics. And once the perceptual theory has been *located* within the texts, there is then the more difficult problem of isolating it from the ontological system in which it is firmly embedded. An added handicap is that in doing this one must

cut through a great deal of terminology and coarse style to behold the theory in its essentials.

A second reason for the lack of widespread discussion of Bergmann's thought is that Bergmann himself has tended to be out of sympathy with the prevailing methods and problem-areas of recent discussion in the philosophy of mind. One of my goals is to show that Bergmann is not as far removed from the main currents of recent discussion as one might think. Through patient exposition and careful criticism, I shall try to show where there is common ground between Bergmann and other strands in the philosophy of perception, and also where I believe Bergmann has made valuable contributions to the current debate.

I shall begin with a sketch of Bergmann's concept of *veridical perception*. This first section will be largely expository and foundational, setting the stage for later discussion. Its initial aim is to acquaint the reader with "the act" and its mental properties, but it also includes a short discussion of Bergmann's conception of the *objects* of perceptual acts. In the following sections, I shall present Bergmann's views on *perceptual error* and *appearing*, and on *sensations*, and work from them toward what I believe to be a more adequate understanding of these concepts.

#### VERIDICAL PERCEPTION: (1) THE MENTAL ACT

The fundamental idea underlying Bergmann's account of veridical perceptions is that they should be viewed as a species of mental act (thought) intending nonmental and actual states of affairs, also called "facts." Thoughts, i.e., individual thinkings, are construed as (bare) particulars exemplifying two "simple" mental properties: a propositional attitude (Int, 7, 28; Acts, 36)<sup>3</sup> and

<sup>1</sup> Gustav Bergmann, "Intentionality" in *Meaning and Existence* (Madison, 1960), pp. 3-38; "Acts" and "Realistic Postscript" occur in *Logic and Reality* (Madison, 1964), pp. 3-44 and 302-340 respectively. Hereafter, references to these three essays will be included in the body of the text in the following abbreviated fashion: (Int, 3), (Acts, 4), (RP, 302).

<sup>2</sup> The general theory of acts, couched in a frame of intentional realism, worked out in these essays, is also maintained in Bergmann's more recent book, *Realism* (Madison, 1967).

<sup>3</sup> Bergmann refers to the "species" of an act, or to its "mode of awareness," rather than its "propositional attitude."

what Bergmann refers to as a "propositional character" (Int, 28). In holding that mental acts exhibit these two nonrelational properties, Bergmann is rejecting G. E. Moore's claim that mental acts are diaphanous.

When Bergmann says that mental acts have propositional characters, he is committing himself to the idea that propositions or judgments enter into perceptions. He refers to these unique propositional properties of acts by means of a quoting device which forms new predicates out of sentences: thus he would speak of a "this-is-green" mental act.<sup>4</sup> Since the intentionality of thoughts is *not* one of the topics of this essay, I shall have very little to say about these unique intentional properties. I shall simply assume that individual thought-episodes do have a propositional property, and I shall refer to this property by means of Bergmann's special predicates (e.g., "this-is-green").

Now let us turn to the other simple property of acts, the propositional attitude. Bergmann holds that a thought always has the property of being, e.g., a perceiving, a believing, a thinking of, a doubting whether, a sensing, etc.<sup>5</sup> It should be noted that in the case of perceiving, Bergmann allows that the mental property in question may be the specific types of perceiving: visual, tactile, etc. (Acts, 36). He thinks it is necessary to introduce these mental properties to account for certain introspectible differences among mental acts which have the same propositional character. Now I do believe that Bergmann is on good ground in making such a claim, at least in regard to perception. For I do not see how it can be denied that there are noticeable differences, which anyone can detect and report about, between the mental experiences as experiences of *perceiving* that something is green and merely *thinking* that it is green. I shall take it that Bergmann's mental property, *being a perceiving*, refers to this introspectible *perceptual character* of the former kind of experience. However, I shall later call into question the idea that this perceptual character of some mental states is "simple," if this means that it cannot be defined in terms of other mental concepts.

Now if, as Bergmann says, *being a perceiving* describes a nonrelational property of a mental act,

then it dawns immediately that his conception of perceiving (and hence of seeing) differs radically from that of most contemporary philosophers. For it is generally accepted that the ordinary concept of seeing is one which describes a *knowledge*-relation between an observer and an actual object; seeing—that is taken to be a specific kind of knowing—that. Bergmann, however, uses "see" in such a way that

1. *S* sees that this is green

does not imply that this is green. The concept of seeing pertains solely to a mental act, and does not carry any implication about the truth of the proposition involved in that act, nor about the grounds or justification for accepting that proposition.

Now Bergmann might argue, in defense of this usage, that there is some basis in common speech for construing the notion of seeing in the way that he does, and not as a type of knowing. It is natural for someone to say, for example, that he had *seen* a ghost, or an unidentified flying object, even when he has good grounds for believing, or even knows, that what was seen in these cases did not exist.

However, granting that these are correct uses of "see," it is still questionable whether the perceptual verb in these contexts is being used to describe a property of a mental state as Bergmann suggests. For it should be noted that in the above contexts, it would be equally appropriate for the subject to report what he *seemed* to see, or what *appeared* to him ("There appeared to be a ghost there"). But I shall later interpret Bergmann as saying that these seems-statements and appears-statements do not function solely to describe a mental property of a mental state, but *also* to call into question the truth or falsity of that mental state. Consequently, it seems that Bergmann's use of "see" and "perceive" to refer to an introspectible feature of a mental state is a technical use of these expressions without any clear basis in common speech. Of course there is nothing wrong with using ordinary terms in special ways, as long as, when one is done, things hang together in the way they are supposed to.

One interesting consequence of Bergmann's use of "see" is that it becomes correct to say that cases of both veridical and nonveridical perception are

<sup>4</sup> I have added hyphens to Bergmann's single quotes to make these unique predicates more conspicuous to the reader. These predicates may be viewed as technical expressions whose ordinary language equivalents are the that-clauses which occur in such ordinary contexts as: "George's thinking that this is green occurred at *t*."

<sup>5</sup> I take it that the same mental act may instantiate *more than one* attitude, although Bergmann does not, to my knowledge, deal with this question. It seems we must allow that the very same act which is a perceiving that this is a chair may also be a believing that this is a chair, which for Bergmann is a distinct kind of act.

cases of *seeing*. For since the mental states, considered as mental states, that occur in *veridical* perception could well be identical in kind with those that occur in *nonveridical* perception (if the subject is deceived), then both would exhibit the property of being a seeing in Bergmann's sense. Now I pointed out above that I believe it is necessary to speak of some mental acts as having a *perceptual character* in order to distinguish them from other nonperceptual acts. And certainly the acts involved in appearances exhibit this same perceptual quality. However, I shall argue that there is a profounder reason for grouping appearances and veridical seeings under the common head of perceiving as Bergmann does. In doing so, I shall defend the view, which differs from Bergmann's, that the concept of a *perceiving as a knowing* is the basic concept in our ordinary beliefs about perception, and that the concept of an appearing presupposes it.

#### VERIDICAL PERCEPTION:

##### (2) THE PERCEPTUAL OBJECT

Bergmann's account of the objects of ordinary perceptions turns on a key distinction between two types of perceiving (1 and 2) that differ in respect of their intentional objects. The intended objects of perceiving<sub>2</sub>, Bergmann says, are such "ordinary things" as chairs and tables. Hence the propositional characters of these perceptual acts are about ordinary things, e.g., "this-is-a-table." A perceiving<sub>1</sub>, on the other hand, intends (in the case of vision) an area-particular (RP, 319) rather than a whole object. Bergmann introduces the notion of a perceiving<sub>1</sub> and its object by making a phenomenological point about perceiving<sub>2</sub>:

When I perceive<sub>2</sub> the table, my "attention does not center" on any of the particulars "in" it that may or may not on this occasion be presented to me.

He continues:

But we all are, when perceiving<sub>2</sub>, capable of shifting to another act whose intention is the fact of some particular (or particulars) exemplifying some properties (and relations). These particulars are all "in" the table; . . . The species of [this act] is again perceiving. . . . Yet [its intention] is characteristically "narrower" than that of the preceding perceiving<sub>2</sub>. An act with such an intention I call a perceiving<sub>1</sub>. (RP, 314.)

Here I take Bergmann to be saying, correctly, that while observing a table, we are capable of shifting to another perceptual act in which what is

seen is a fact about some area-particular. One might achieve this state by focusing more narrowly on the facing surface of the table. The propositional character of such an induced minimal perceiving would be expressed by a sentence of the form

This area-particular has such-and-such properties.

However, it should be apparent that Bergmann in the above passage is making a much stronger claim than this phenomenological one. He is in fact maintaining that the intentional objects of perceiving<sub>2</sub> can be *analyzed* in terms of a set of particulars having observable properties. He goes on to say that what has been said about perceiving<sub>1</sub> "provides a cue for assaying the intentions of perceiving<sub>2</sub>," and then proceeds to analyze ordinary objects (or rather facts pertaining to them) as follows. (Here I must include a rather lengthy and tortured passage, as it introduces terminology without which the reader will be lost in later discussion.)

Let '*Mp*' be the sentence representing the fact that would have been presented to me if the act had been a perceiving<sub>1</sub> instead of a perceiving<sub>2</sub>. Let '*a*<sub>1</sub>, . . . , *a*<sub>*n*</sub>' stand for the particulars "in" it and, whenever it helps, write '*Mp*(*a*<sub>1</sub>, . . . , *a*<sub>*n*</sub>)' instead of '*Mp*'. The letter '*M*' in '*Mp*' is to remind us that the fact is molecular. The letter '*p*' is to recall 'part'. For I assay *P* (This is a table), which is the intention of the perceiving<sub>2</sub> with which we are concerned, as a conjunction of two (part) facts. One is *Mp*; the other I call *Op*. That makes '*P*' an abbreviation for '*Mp*·*Op*'. The letter '*O*' is to remind us that '*Op*' contains operators. *Op*, the operator part of *P*, is very complex indeed. Fortunately we need no details. The idea is easily grasped. '*Op*' schematically states that there are all the particulars which must be there, that these particulars have all the properties they must have, and that they stand in all the relations in which they must stand, among themselves and to the particulars in the molecular part, if the latter is to be "in" a table. (RP, 314-315.)

In this passage, Bergmann is putting forward the improbable idea that the area-particular(s) which I *would* perceive, if I were perceiving<sub>1</sub> (*Mp*), is the core ingredient of what I *do* perceive when I perceive<sub>2</sub> the whole table. This claim carries with it the idea that the only particulars which are "present" to me when I perceive a table are those which constitute its facing surface.

Now I believe that Bergmann gets on the wrong track in attempting to analyze the ordinary objects of perceiving<sub>2</sub> in terms of the objects of the

corresponding perceivings<sub>1</sub>. For it seems to me a mistake (although I shall not try to prove that it is) to think that the area-particulars seen in perceivings<sub>1</sub> are in any sense *parts* or *constituents* of ordinary objects. The area-particulars which Bergmann thinks are seen in normal perceptions are actually very special objects seen in very special circumstances (usually experimental situations). Hence Bergmann's error does not lie in thinking that area-particulars can be perceptual objects, but in taking them as the "cue" to the perception of ordinary objects.

In the remainder of this essay, I shall be concerned primarily with instances of perceiving<sub>2</sub>, where the intentional objects are ordinary things. I shall argue that the notion of a perceiving<sub>1</sub> is a useful one only because of the role it plays in helping us to understand the concept of a sensation.

### PERCEPTUAL ERROR

For Bergmann, there are two kinds of perceptual error: *qualitative* and *existential*. In the former, some actual thing appears to be other than it is, whereas in the latter, there appears to be something which does not actually exist, as in hallucinations. I shall limit my discussion to Bergmann's concept of qualitative appearing.

We have seen that cases of perceptual error are to be construed as *perceivings* in Bergmann's sense, and that they could involve the same propositions as veridical perceivings. The difference between veridical and nonveridical mental acts is that the proposition involved in the former is true, and in the latter false.

Bergmann's main concern is with the ontological implications of this epistemic difference. He claims that in cases of error where we are not related to an actual state of affairs, we are nevertheless still related to something, namely a *non-actual* state of affairs, or as he puts it, somewhat paradoxically, a "possible fact." The essential point is put this way: "In assaying [qualitative

error] there is only one hurdle to overcome. One must recognize that possibilities exist" (RP, 321).

Let us look at how Bergmann explains this point about qualitative appearing in a particular case. We should note that his entire discussion of error turns on the preceding analysis of perceivings<sub>2</sub>. The example given is the following:

I perceive<sub>2</sub> an oval coin. Surprised that there should be such a thing, I reach for it, examine it, perceive<sub>3</sub><sup>6</sup> it to be round. The '*Op*' of the original perceiving<sub>2</sub> is false. How about its '*Mp*'? Let *a* be the particular (area) "in" it that was presented to me. '*a* is round', I have now reason to believe, is true; '*a* is oval', false. The latter is a conjunction term of '*Mp*'. That makes *Mp*, too, a mere possibility. But the *external particular* in *Mp* is *real*. (RP, 319.)

Though we have already been introduced to the terminology in this passage, it is nevertheless so compressed that a commentary is necessary. I offer the following construction of what is being said.

In this example, the perceptual judgment involved in the original perceiving<sub>2</sub> is expressed by the false sentence "this is an oval coin." In making this judgment, the perceiver is intending something about all the particulars in the coin, including of course the area-particular in the facing surface. Now in the above passage, Bergmann suggests that the perceiver, when judging<sub>2</sub> that this is an oval coin, is at the same time making a *second* judgment, which corresponds to *Mp* in the object, namely the judgment that *a* is oval. Hence he seems to be assuming, tacitly, that when one judges that something is an oval coin, he is committed to the further judgment that the particular in the facing surface of the coin is oval.<sup>7</sup>

On this interpretation, Bergmann's point comes to this: since both of the judgments involved in this perceptual experience are false, then their intentional objects are merely possible states of affairs. However, the individual thing which these possible states of affairs concern is actual. The coin is real!<sup>8</sup>

Clearly the crux of Bergmann's analysis of perceptual error lies in his ontological claim about

<sup>6</sup> This is not a typographical error. A perceiving<sub>3</sub> for Bergmann is the final perception<sub>2</sub> in a series of perceivings<sub>2</sub> aimed at checking the truth of the first perceiving<sub>2</sub> in that series. The judgment involved in a perceiving<sub>3</sub> is therefore a justified one. (RP, 316.)

<sup>7</sup> The situation is really more complicated than Bergmann admits. I am speaking as if *Mp* is how I visually *take* the facing side of an object to be when I see (or seem to see) that object. But this is not necessarily how I think the facing side of the object *appears*. And consequently, neither is it how the area-particular which I would perceive<sub>1</sub> if I abstracted the facing surface from the object would appear. It is really all three of these things that Bergmann is calling *Mp*.

<sup>8</sup> The account Bergmann gives of existential appearances (RP, 320-321) leads to a similar conclusion, with the important difference that the individuals which the intended possible states of affairs concern are *not* real. If I perceive a ghost, the nonactual ghostly fact which is the intentional object of this perception concerns a nonactual individual ghost. If we accept the idea of nonactual facts, then I see no reason to balk at nonactual individuals.

possible states of affairs. Although it is not my purpose in this essay to determine the ultimate status of these intentional objects, I shall conclude this section by gathering together the most important claims Bergmann makes about them, and commenting briefly on each.

(1) Possible states of affairs have ontological status, though they are not actual. (2) They are what mental acts are related to in nonveridical perception. (3) In perceptual cases, they are rich, sensuous states of affairs, which may be indistinguishable from the perceptible facts constituting ordinary objects.

Bergmann is on firm ground, I believe, in making each of these claims. I am in agreement with (1) and (2) because it seems that if one accepts the idea that mental acts have a propositional or judgmental character, as I do, then one is automatically committed to the intentional objects which those judgments are about, and to which one must make some sort of existential reference. Bergmann reminds us that we must distinguish the act and its object, the *perceiving* and the *perceived*. If I seem to see a ghost at my window, that which I seem to see, this ghost, is certainly distinct from the act or perceptual experience which intends that object and through which I see it. Moreover, concerning (3), the claim that the intentional objects of perceiving have a "sensuous" character, it is difficult to deny that the states of affairs intended by perceptual acts, veridical and nonveridical, are qualitatively different from those intended by non-perceptual acts such as mere thinkings.

It seems that we must accept, provisionally at least, the conclusion that the intentional facts of which Bergmann speaks exist in some sense, though they are not actual.

#### APPEARING

Bergmann's primary concern, in his discussion of perceptual error, was to arrive at an adequate *ontology* of appearing. The burden of his analysis of cases of nonveridical perception was to show that in these cases the existence of nonactual states of affairs must be admitted. In this section, I shall move beyond Bergmann's rather circumscribed ontological concerns and ask whether or not some general theory or definition of the *meaning* of *appears-statements* can be elicited from Bergmann's account.

One serious limitation of that account is that it is restricted to cases of perceptual *error*, where there

appears to be something which does not actually exist. Although I shall initially work within this framework of *mere appearances*, my long-range goal is to construct, if possible, a more general theory of appearing which will have application also to cases in which what appears to be the case *is* the case.

Before beginning this project, it will be helpful to review in a general way what Bergmann has said about appearing. An important feature of his account is that he has tended to assimilate appearances and their objects to veridical perceiving and their objects. He has done this by emphasizing that the mental states, or more exactly acts, involved in these two kinds of experiences could be identical, that is, they could exhibit the same propositional and perceptual character. In addition, he has said that the objects of these two experiences, viz., states of affairs, might be the *same* states of affairs in that they might be perceptually indistinguishable, and expressed by the same statement, although they differ in respect of their "mode" of existence. In the remainder of this section I shall explore this assimilation of nonveridical to veridical perceiving and also argue that it must be strengthened in a way that Bergmann himself would not allow.

The following general picture of the meaning of *appears-statements* seems to emerge from Bergmann's discussion. If we say of George:

2. There (merely) appears to George to be an oval coin before him,

then we are asserting two things: (a) we are describing George's occurrent visual experience as being a perceiving that there is an oval coin before him (where "perceiving" in this context has Bergmann's technical, nonepistemic sense in which it describes a mental property). More exactly, we are characterizing George's experience as being a perceptual one, and as having a certain specific propositional character ("this-is-an-oval-coin"). In addition, (b) we are making an epistemic judgment that the proposition involved in this experience is false and hence that George (his act) is not related to an actual state of affairs.

And if George were to make a corresponding report about his own experience:

3. There (merely) appears to me to be an oval coin over there,

he too would be describing the kind of experience he was having, as well as asserting its falsity.

On this showing, then, statements about mere

appearings have both a *descriptive* and an *epistemic* function: they describe the subject's occurrent experience, and also deny the truth of the propositional content of that experience.

It may help to clarify this interpretation of Bergmann if we express it in terms of the ontological framework of mental acts. When George reports, as in (3), that there merely appears to be an oval coin before him, Bergmann seems to be saying that his mental state includes two distinct acts. There is, first, George's *perceiving* (in Bergmann's sense) that the coin is oval, and secondly, his *judging* about this perceiving that it is false, i.e., that its object is a possible fact.

Now I believe that this account of mere appearings, which I am attributing to Bergmann as being very much in the spirit of his remarks, does provide us with all but one of the essential insights needed to construct an adequate and general theory of appearing. First of all, Bergmann is surely right in holding that appears-statements have the descriptive role he gives them. When someone reports in a perceptual situation that something *x* appears to have some property *f*, then whatever else he might be saying, he is at least reporting his occurrent experience. Moreover, it is correct to say that the proposition that he ascribes to his experience has the form "*x is f*" (rather than "*x appears f*"). And finally, I think Bergmann is right in linking the concept of *appearing* with the concepts of *truth* and *falsity*, though as we shall see, this connection is somewhat more complicated than Bergmann recognizes. However, the fundamental mistake of this account, as I shall argue in a moment, is that it fails to recognize a more intimate connection that exists between *appearing*, in general, and *veridical seeing*.

At this point in the discussion, if we are to achieve our goal of finding a general theory of appearing, we must find some way of dealing with the fact that not all appearings are *mere* appearings, and not all appears-statements imply that their constituent propositions are false. Thus someone might say, while looking at a green wall in normal conditions:

4. This wall appears green to me, and of course it is green.

In this case, the proposition that the wall is green is *true*.

Moreover, as this example shows, we cannot say that it is at least always part of the meaning of appears-statements that they *call into question* the

truth of their constituent propositions; for although appears-statements are often *used* to express doubt about what appears, this cannot be what these statements *mean*, since they can be correctly used even when the speaker has no doubt whatsoever about the truth of their propositions. The general point seems to be that if one asserts that *x appears f*, he does not thereby commit himself to either the truth or the falsity of the proposition that *x is f*.

We may now ask: is it possible to give an account of appearing which is consistent with this point, but at the same time retains Bergmann's insight that the concept of appearing is an epistemic one involving in some way the notion of truth? The alternative would seem to be to abandon the link between *appearing* and *truth*, and hold that there is at least one important sense of "appears" in which it has a purely descriptive role, i.e., simply describes the subject's occurrent experience.

Now I do not see that we are driven to this alternative, for I hold that we can, and should, explicate the concept of appearing as an epistemic one, by means of the concept of *veridically perceiving*. The basic idea of the view I wish to defend is that when someone judges how something appears he is assimilating the perceptual state he is in to a corresponding state of actually perceiving. However, it is crucial to realize that the concept of perceiving which I am employing is not Bergmann's concept of perceiving (i.e., it does not name a simple mental property), nor is it even what Bergmann would refer to as veridical seeing (i.e., it does not simply describe a relation between a mental act and an *actual* object). Rather *perceiving*, on my account, has the sense of *knowing*, and carries with it the implication that the proposition involved in the perceiving is not only true, but justifiably accepted. In attempting to understand appearings in terms of the notion of *perceiving as knowing*, I am departing fundamentally from Bergmann's view.

I stated in the last paragraph that when someone makes an appears-judgment he assimilates his experience to that of actually perceiving. The judgment involves, as it did for Bergmann, both a descriptive and an epistemic claim: (a) the subject is *comparing* his experience, as an experience, in both its propositional and perceptual character, with the experience of actually perceiving the thing in question. At the same time, (b) he is *checking* or *canceling* the normal truth-implication of that perceiving.

To apply this claim to our notorious case of the

oval coin: when someone reports that a coin appears oval to him, he says, in effect,

this experience of mine, as an experience, is the same as (or closely resembles) the experience involved in actually seeing an oval coin, though I do not (for whatever reasons) commit myself to the idea that the proposition involved in that experience is true (as I would if I were actually seeing an oval coin).<sup>9</sup>

Now with regard to the epistemic part of the meaning of appears-statements, (b) above, it should be realized that the *checking* of the truth-implication of one's perceptual state does not imply that the subject believes or even suspects that his perception is or may be false. It is indeed true that very often when someone says how something appears it is because he has become aware of some unfavorable evidence which casts doubt on his perception. However, we have seen that it cannot be part of the meaning of appears-statements to call into question in this way the truth of a perception. All that can be said, and all that needs to be said, is that in judging how something appears, one does not by that judgment commit himself to either the truth or the falsity of his perception. And this is perfectly consistent with his making another overriding judgment in that situation that the propositional content of his experience is indeed true, or indeed false. What is important is that the appears-judgment cancels the truth of the perception, which automatically marks it as an essentially epistemic judgment.<sup>10</sup>

As to the descriptive part of the meaning of appears-statements, (a) above, it can be seen that I agree with Bergmann that in making these statements one is ascribing a certain propositional character and a perceptual character to his experience. However, I have argued that one does not do this, as it were, directly, but indirectly by applying to his experience the concept of a corresponding seeing which has the characteristics in question. I would also agree with Bergmann that appearances involve two acts rather than one inasmuch as an appearing includes a reference to a corresponding perceiving which is, so to speak, the

core act of the appearing, and which is qualified in the complex way described above.

A final observation. The idea that appearances should be analyzed in terms of veridical perceivings finds support in the fact of ordinary usage that sentences of the form "*x looks f (to me now)*" are usually synonymous with "*it is as though I were seeing x to be f.*" The account of appearing I have given secures the conceptual dependence suggested by this usage.

What is most important is to retain the idea that the concepts of appearing and perceiving are essentially bound up with those of knowledge, truth, and evidence. Bergmann is in danger of losing sight of this fact in failing to recognize the conceptual connections that I have stressed.

#### SENSATIONS: BERGMANN'S THEORY

In this section I shall explore Bergmann's concept of sensations and try to determine grounds for deciding whether or not this concept is an acceptable one. The broad questions I shall put to Bergmann are three: (1) what grounds are there for thinking that sensations exist? (2) how do we go about characterizing sensations? and (3) what role do sensations play in perception? But before we attempt to decipher Bergmann's claims about sensations, it will be helpful to reflect for a moment on the general framework into which sensations must fit, and specifically, to ask whether the analyses we have given of *perceiving* and *appearing* place limitations on the possible meanings that can be given to the concept of sensations.

In the first place, the *realistic* assumption of our analysis of perceiving, which rests on a fundamental dichotomy between the subject's act and its actual physical object, places definite restrictions on what we can say about sensations. Specifically, within this realist framework, we can assume that sensations, or sense objects,<sup>11</sup> are not identical with, or constituents of, the ordinary objects of perceptual acts, such as tables and trees. If sense objects do play a role in veridical perception, then it is clear that they must be "located" in the subject's mental

<sup>9</sup> The account of appearing that I have developed is based on the account given by Wilfrid Sellars in "Empiricism and the Philosophy of Mind" in *Science, Perception and Reality* (New York, 1963), sec. III, pp. 140-149, and differs only in minor respects from it.

<sup>10</sup> Sentences such as (4) above, "This wall appears green to me, and of course it is green," sound odd because one who uses the sentence is granting in the second clause what the first clause suggests he refuses to grant. Cf. "Mary will probably arrive tonight, and in fact Mary will arrive tonight."

<sup>11</sup> I am here anticipating Bergmann in speaking of sense objects (Bergmann uses the term "sense data"). For the moment, we may think of them as the entities that are sensed in a sensation.



experience as an experience. And to be sure, Bergmann holds that sense objects are *mental* entities in the sense that they are mind-dependent (exist only so long as the subject is having them) and *private* (can only be directly known by the subject who has them).

Moreover, when we consider how sensations might figure in *nonveridical* perception, we arrive at parallel conclusions. For we interpreted mere appearances as mental acts intending perceptual states of affairs: when there appears to be a green spot before me, then my (false) perceptual judgment ("this is a green spot") is intentionally related to a nonactual state of affairs (this being a green spot). And there does not seem to be any justification for identifying the perceptual fact of something's being a green spot with an object of sensation. At least Bergmann does not make any such identification.

How exactly do sensations function in the perceptual experience? This is a question on which our analyses of appearing and perceiving have so far shed very little light. We saw that appears-statements do describe one's experience, but only indirectly, by assimilating it to the experience involved in a corresponding perceiving. And although we did distinguish the propositional and nonpropositional components of the perceptual experience, we have not yet found any way of directly characterizing the nonpropositional component except to say, with Bergmann, that perceptual acts exhibit the property *being a perceiving*. But while saying this does mark the difference between perceptual and nonperceptual acts such as thinking and believing, it does not show us any way of characterizing the *specific features* of the nonpropositional component of perceptual experience. I shall argue below that we can describe these specific features by coming to a proper understanding of sensations and their properties.

We turn now to Bergmann's account of sensations as he presents it in the two essays, "Acts" and "Realistic Postscript." Bergmann bases his argument for sensations on certain facts of introspection; he holds that sensations can be discovered in one's mental state (or as he says, "conscious state") by performing a mental shift. What one notices while in this special mental set are certain colorful entities or "sense data" existing momentarily in one's conscious state. However, the role that these sense data play in perception is obscure in Bergmann's theory. The crucial claim is that sense data do not figure in the analysis of perceivings: perceivings can be fully explicated in terms of mental

acts and their intentional objects. Because of this, Bergmann concludes that "dialectically it makes no difference whatsoever whether or not there are sense data" (RP, 325, italics omitted).

In presenting his idea of sense data, Bergmann uses his earlier distinction between perceiving<sub>1</sub> and 2. To refresh the reader's memory, we said that a perceiving<sub>2</sub> intends a whole physical object while a perceiving<sub>1</sub> intends only the constituent particulars which can be more narrowly focused on—in the case of vision, the area-particulars "in" the facing surface. In addition, Bergmann makes a key use of the notion of *externality*, which he takes to be a characteristic of all perceived<sub>2</sub> and perceived<sub>1</sub> objects, but not of sensed objects. Here are the relevant passages:

Everyone can make the shift from perceiving<sub>2</sub> to perceiving<sub>1</sub>. . . . The molecular fact intended by a perceiving<sub>1</sub> (of mine) still presents the "idea of external existence." . . . Whatever else may or may not be "in" [my conscious state], the molecular fact, its intention [*Mp*], certainly is not. (RP, 325.)

He continues:

To claim that there are sense data is to claim that, just as we can shift from perceiving<sub>2</sub> to perceiving<sub>1</sub>, we can make a second shift, in the same direction, as it were, from perceiving<sub>1</sub>. . . . If the intention from which you shifted is *Mp*, call this conscious state  $\overline{Mp}$ .  $\overline{Mp}$  is exactly like *Mp*. Only the [externality] has disappeared! This conscious state  $\overline{Mp}$  is a sense datum. (RP, 325.)

The most plausible interpretation of the second shift Bergmann describes, from perceiving<sub>1</sub> to sensing, is that it is a shift from seeing something which is *not* in one's conscious state to seeing (or sensing) something which *is* in one's conscious state. This is why Bergmann stresses the fact that the object of the perceiving<sub>1</sub> is not in one's conscious state. Further, even though the objects of the perceiving<sub>1</sub> and the sensing are "exactly like" one another, Bergmann distinguishes them on the grounds that the perceived object, but not the sensed object, "presents the idea of external existence." How does this notion of the presence or absence of externality shed light on sensations?

While Bergmann takes the "idea of external existence" as presented by an object to be a simple datum of consciousness (RP, 311) it is natural to interpret this notion in terms of the idea that the object that one sees (or seems to see) is *in physical space*. If this is correct, then in claiming that the objects of sensation do not present the idea of

externality, Bergmann would be saying, plausibly, that when one shifts from a perceiving to a sensing the sensed object is no longer spatially separated from the subject in the way that the perceived object was: the sensed object is not (or does not appear to be) in physical space.

Now it is difficult to deny the introspective facts to which Bergmann is calling attention, namely that we can come to notice the visual objects he describes by changing our mental set. Moreover, it seems correct to say that these objects are not seen as being in physical space, viz., they are not in the vicinity of or near the surface of the objects we were looking at before shifting into our phenomenological frame of mind. Hence I conclude that Bergmann has given us some reason for thinking that sense data exist. Furthermore, it would seem to be a justifiable inference that sense data are present in *all* perceptions, since whenever we are perceiving we are capable of noticing them by shifting into the proper mental set.<sup>12</sup> But still, it is not at all clear what *role*, if any, sense data play in perception, and how they are connected with perceptual acts. Specifically, we may wonder whether it is any longer necessary to say that perceptual acts have the simple property *perceiving*, once it is recognized that sense data always accompany these acts in one's conscious state.

#### SENSATIONS AS COMPONENTS OF THE ACT

My objections to Bergmann's conception of sense data are two. First, it obscures the role of sensations in perception and ultimately leads to the idea that they are irrelevant to a philosophical understanding of ordinary perceiving, as happens with Bergmann. Secondly, it loses sight of an essential feature of the perceptual experience as an experience, namely that it has the same kind of sensory properties that Bergmann ascribes to sense data. What I am suggesting is that it is the perceptual experience as a whole which may be said to have sensory properties (as well as a propositional character).

However, this claim must be qualified in an

important respect. We must take into account something which Bergmann overlooks, namely that mental entities, of whatever sort, are *not* things that can have sensory properties. A sense datum, for example, cannot be literally green and triangular since these properties pertain primarily to physical objects in space. Hence what must be said about inner experiences is that they have properties (let us call them *phenomenal* properties) which are different from, but *analogous* to, physical sensory properties such as color and shape.<sup>13</sup> It is these phenomenal properties which, I am arguing, may be ascribed to the perceptual experience. If I am right about this, then a correct description of the experience that occurs in the perception of a green triangle is that it has, first, the propositional character "this-is-a-green-triangle" (cum attitude), and second, unique sensory properties analogous to greenness and triangularity.

But there is no reason why we should not refer to the sensory or nonpropositional component of the perceptual experience as the subject's *sensation*. That is, we seem to have found a useful and commonsensical concept of a sensation as a component part of the perceptual experience. Indeed, once this suggestion is taken seriously, then it dawns on us that the introspective evidence pertaining to sense data that Bergmann brings forth does not justify his inference that sense data and perceptual act are *distinct entities*. For all Bergmann has shown, it is quite possible that when one makes the shift from a perceiving to a sensing, he is coming to notice properties of his perceptual *act*.

I shall try to defend my nonBergmannian concept of a sensation, and specifically the idea that it has the phenomenal features I described, by drawing on one of our conclusions about appearances. I argued earlier that in making appears-reports, one is comparing his experience with that of a corresponding seeing. Moreover, I suggested that the subject is able to recognize the *respects* in which his experience resembles and differs from that related experience. If this was correct, then the report that there appears to be a green triangular object over there involves the claim that

<sup>12</sup> What should be said about Bergmann's perceiving<sub>1</sub>? Has he succeeded in describing an intermediate stage between perceiving whole objects (or rather the facing surfaces of whole objects) and coming to sense sense data—a stage in which one sees an expanse or area as being in external physical space? Perhaps it is necessary to pass through such a stage in coming to notice our inner experiences; but I would argue that whether it is a phenomenological question which, to use a Bergmannian turn of phrase, is dialectically irrelevant to the question of sense data and their characteristics.

<sup>13</sup> Wilfrid Sellars describes the analogy between physical and phenomenal color and shape in terms of common logical and structural features in "The Structure of Knowledge: (1) Perception" (unpublished), p. 20. I would add that these physical and phenomenal properties are *disanalogous* as regards their causal features.

this perceptual experience (of mine) as an experience resembles the experience involved in actually seeing a green triangular object over there *in respect of its greenness and triangularity*.

Now in reporting comparatively about the greenness and triangularity of one's experience, I think it is clear that one is not simply referring to the propositional element of the two experiences (which does include the concepts *green* and *triangular*). One is also characterizing one's experience in its *nonpropositional* aspect by means of the predicates "green" and "triangular." In this way, I believe, our analysis of appearances lends support to our concept of sensations as that part of the perceptual experience having properties which are counterparts of the perceptible properties of their intended objects.<sup>14</sup>

Once we have put this concept of sensations to use in understanding perceivings, then there is no longer any motivation for saying, with Bergmann, that the perceptual character of perceivings is a simple mental property, for we have succeeded in giving a more determinate description of this property in terms of sensations. Moreover, if as I have claimed, sensations are component parts of perceivings, then they are as dialectically relevant in giving an analysis of perceiving as is the act itself.

Finally, it is important to realize that my defense of sensations has not rested on the claim that they are ever observed, i.e., that they are ever perceptual objects. All I have said is that one can recognize certain things about his mental experience, which is simply to say that one can give noninferential reports about his experience. And while I certainly do not deny the relevance of the introspective facts that Bergmann calls attention to, I believe I have shown that Bergmann is wrong in claiming that these facts are the *sole* grounds for introducing sensations, for I have found other common sense considerations which lead to this idea.

## SUMMARY

To conclude this discussion, I shall review briefly some of the more important elements of Bergmann's perceptual theory, as well as where I have found them to be in need of revision. (1) On the important matter of *nonveridical perception*, Bergmann has made a strong case for saying that the intentional objects involved in these perceptions exist in some sense, though they are not actual. These nonactual states of affairs, he pointed out, may be perceptually indistinguishable from actual states of affairs. Although Bergmann's concept of these nonactual states of affairs is not a wholly satisfactory one, he has at least shown us the place they occupy in perception, and hence the road that any rival theory must travel.

(2) Bergmann rightly drew attention to the *perceptual character* of perceptual acts, which distinguishes them from such nonperceptions as merely thinking or believing. However, we found that it was not necessary to stop with the idea that *being a perceiving* is a simple mental property of acts, but that it could be further explicated in terms of a concept of sensations.

(3) In regard to *appearing*, although Bergmann correctly assimilated the concept of appearing to that of veridical perceiving, we found it necessary to strengthen this assimilation in such a way that appearances must be *analyzed* in terms of veridical perceivings (in the sense of knowings).

(4) Concerning *sensations*, Bergmann has stressed the introspective grounds which justify, in part, philosophical claims about sensations, thus helping to secure their place in the philosophy of mind. We supplemented his thesis by constructing a different, but commonsensical, notion of sensations which did not rely on introspective evidence, and which made sensations an integral part of the perceptual process.

University of Pittsburgh

Received February 9, 1973

<sup>14</sup> I should make it clear that I am *not* claiming that the concept of sensations that I have given is the only acceptable one, philosophically speaking, nor that it is the most penetrating one. Specifically, I do not wish to deny that there are good grounds for speaking of sensations as entities (or states) separate from the perceptual experience. In fact, I believe that there are convincing reasons for inferring the existence of antecedent sensory states which mediate between the purely physical state and the "full-blown" perceptual experience. (See Wilfrid Sellars, *Science and Metaphysics* [New York, 1968], ch. 1, for a valuable discussion of what Sellars calls the "sense impression inference.") One such reason is to give an adequate explanation of the sensory component of perception that I have described. But the fact that such theoretical advances are necessary should not blind us to the more rudimentary understanding of sensations as the nonpropositional component of perceivings.

## IV. HOW DO YOU KNOW?

ERNEST SOSA

DESPAIR of knowing what knowledge is dates back to Plato's *Theaetetus*. Most recently, the trinitarian view of knowledge as justified true belief has been refuted, and a multitude of problems has appeared. Progress on this question is perhaps fated to be asymptotic. But such progress as can now be made depends, in my opinion, on a careful study of the conditions within which a correctly believed proposition is a bit of knowledge. In what follows I hope to enhance our knowledge of knowledge by contributing to such a study.

### I

An accepted truth is knowledge only if evident. What then is it for something to be evident? One short answer is this: a proposition is evident to someone provided he is (theoretically) justified in believing it.<sup>1</sup> But under what further circumstances is the truth of a proposition evident to someone? This is our first main question.

To begin with, there are two general situations where it is evident to someone *S* that *p*. First, there is the situation where it is self-evident to *S* that *p*, i.e., where from the fact that *S* correctly believes that *p* we may infer that it is evident to *S* that *p*. Our inference here cannot be logically valid as it stands, however, since logic alone will not enable us to infer that anything is evident just from the fact that it is correctly believed. Some extralogical principles must be invoked to invalidate such an inference.

Such principles are clearly needed, in any case, provided we want to account for (and not deny) our empirical knowledge. Accordingly, they have

long been recognized and accepted. The Greek Sceptics perceived the problems involved clearly and made some definite suggestions.<sup>2</sup> Descartes and Hume were in fundamental agreement about epistemic principles: There is first the self-evident, which includes some obvious general truths and some particular claims, mainly about one's subjective states; anything else that is evident must be made so by being deduced from the self-evident. This much, it seems to me, they both accepted, even if it led them in very different directions: Descartes to his baroque system, and Hume to his desert landscape. Coming to more recent philosophy, the principles underlying the "criteria" of Wittgenstein and others seem best understood as epistemic principles.<sup>3</sup> Discussion of epistemic principles and epistemic scales is to be found in Chisholm's writings.<sup>4</sup> Finally, the third of Sellars' Matchette Lectures, delivered at the University of Texas, is a discussion of "Epistemic Principles."<sup>5</sup> (Among the correlates of such principles in Sellars' philosophy are his principles of "trans-level inference."<sup>6</sup>)

What would such epistemic principles look like? What would be some examples? Let us introduce the schematic letter '*p*' to be replaced by declarative clauses. The following might then be a correct epistemic principle schema: "If *S* correctly believes that it occurs to him that *p*, then it is evident to *S* that it occurs to him that *p*."

Other examples of correct principle schemata yielding self-evident propositions are those corresponding to basic a priori truths, such as "If *S* believes that either *p* or it is false that *p*, then it is evident to *S* that either *p* or it is false that *p*." Here

<sup>1</sup> The parenthetical qualification is meant to rule out as irrelevant whatever practical reasons there might be for having a belief, such as those of a sick man whose belief that he will recover is essential to his recovery.

<sup>2</sup> Thus the theory of Carneades of Cyrene is presented by Sextus Empiricus in his *Outline of Pyrrhonism* and in his *Against the Logicians*. This theory is discussed by Roderick M. Chisholm in his *Theory of Knowledge* (Englewood Cliffs, 1966), pp. 41-44.

<sup>3</sup> Cf. Norman Malcolm, *Knowledge and Certainty* (Englewood Cliffs, 1963), pp. 113-117.

<sup>4</sup> See *Perceiving: A Philosophical Study* (Ithaca, 1957); and *Theory of Knowledge* (Englewood Cliffs, N.J., 1966). For more recent developments see "On the Nature of Empirical Evidence," in *Experience and Theory*, ed. by L. Foster and J. W. Swanson (Amherst, 1970).

<sup>5</sup> The lectures were delivered in 1970 but are as yet unpublished.

<sup>6</sup> See *Science, Perception, and Reality* (London, 1963), p. 88. Trans-level inference is also discussed in "Empiricism and the Philosophy of Mind," in *Minnesota Studies in the Philosophy of Science*, Vol. I, ed. by Herbert Feigl and Michael Scriven (Minneapolis, 1956). This essay is also Chap. 5 of *Science, Perception, and Reality*.

again from  $S$ 's mere correct belief of a proposition that  $p$  (e.g., that either it is snowing or it is false that it is snowing) we may infer that it is evident to  $S$  that  $p$ , and hence such propositions count as self-evident for  $S$ . Moreover, as distinguished from the group about propositional attitudes the present group of propositions are self-evident-if-evident for every person. That either it is snowing or it is not snowing is self-evident-if-evident for everyone, but it is self-evident-if-evident only to Tom that it occurs to him that snow is white since it is true only of Tom that if he correctly believes that proposition, we can infer that it is evident to him.

We have seen how the first general situation where it is evident to  $S$  that  $p$  is one where it is self-evident to  $S$  that  $p$ . But obviously there are known facts that are not self-evident. Hence if a fact can't be known without being evident, there must be facts that are evident but not self-evident. And, more generally, there are propositions that are evident but not self-evident. This is of course the *second* general situation where it is evident to someone  $S$  that  $p$ . Now when it is evident to  $S$  that  $p$  but not self-evident to  $S$  that  $p$ , then that  $S$  correctly believes that  $p$  is not sufficient to establish it as evident to  $S$  that  $p$ , even with the help of the epistemic principles. What then is thus sufficient?

The answer I wish to propose is, in first approximation, that in such circumstances what makes it evident to  $S$  that  $p$  is a non-empty set of propositions  $\alpha$  such that  $\alpha$  validates the proposition that it is evident to  $S$  that  $p$ . Vaguely put, the idea is simply that if you know that  $p$ , but it is not self-evident to you that  $p$ , then you must have grounds for believing that  $p$ , and these grounds must make it evident to you that  $p$ . But to give your grounds for believing that  $p$  is not necessarily to give a complete epistemic explanation of how it comes to be evident to you that  $p$ . For often grounds that you have for believing that  $p$  cannot be groundless but require grounds of their own, and these may require grounds in turn, and so on. A complete epistemic explanation will not stop until it has adduced the grounds for every ground that has grounds.

It therefore appears that we need more than just the idea of a set of grounds that grounds the proposition that  $p$  for  $S$ . We need the idea of a set of grounds that *fully* grounds the proposition that  $p$  for  $S$ , the idea of a set that gives a complete epistemic explanation of how it comes to be evident to  $S$  that  $p$ .

But how are we to define this idea of a set's fully grounding a proposition for a subject at a time? Shall we say that  $\alpha$  fully grounds  $x$  (for  $S$  at  $t$ ) when

for every ground  $y$  in  $\alpha$  there is a subset  $\alpha'$  of  $\alpha$  such that  $\alpha'$  grounds  $y$ ? But what if  $\alpha = \{\text{that the triangular card is approximately equilateral, that the triangular card is approximately equiangular}\}$  and  $x = \text{that each of the card's angles has approximately } 60^\circ$ ? This appears to fulfill the conditions specified and yet  $\alpha$  can hardly be said to provide a complete epistemic explanation for how it comes to be evident that each of the card's angles has approximately  $60^\circ$ . For instance, how do we know that there is a triangular card at all, or that it is approximately equilateral in the first place?

Adapting an idea of Frege's, shall we say that  $\alpha$  fully grounds  $x$  iff  $\alpha$  is the class of ancestral grounds of  $x$ ? (Here  $y$  is an ancestral ground of  $x$  iff  $(\forall \alpha)\{[x \in \alpha \ \& \ (\forall z)(\forall w)[(w \in \alpha) \ \& \ (z \text{ is a ground of } w) \supset (z \in \alpha)] \supset y \in \alpha]\}$ ; i.e.,  $y$  is an ancestral ground of  $x$  iff  $y$  is a member of every class containing  $x$  and all grounds of members.)

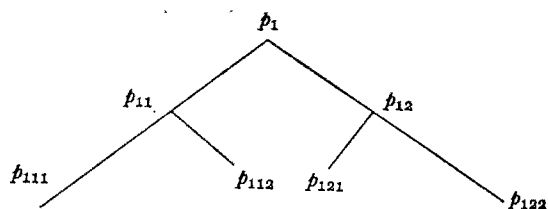
The main problem with this definition derives from cases of epistemic over-determination. In such cases the class of ancestral grounds of  $x$  will contain superfluous grounds. This will turn out to be unfavorable to our purposes, mainly because a defective ground—say a false ground—surely need not flaw every fully grounding class, not if it's a superfluous ground. And yet, according to the Fregean idea, there is only *one* fully grounding class, and it *will* contain *every* ancestral ground, superfluous or not.

My suggestion is therefore this:  $\alpha$  fully grounds  $x$  for  $S$  at  $t$  if and only if there is a set of subsets of  $\alpha$  which form a sequence  $C_1, \dots, C_n$ , such that (i)  $C_1$  grounds  $x$ , (ii) for each member  $y$  of any subset  $C_i$ , such that  $y$  requires grounds, the successor of  $C_i$ ,  $C_{i+1}$ , grounds  $y$ , and (iii) the last subset,  $C_n$ , contains no ground that requires grounds.

We are finally in a position to answer our first main question. A proposition  $x$  is evident to someone  $S$  in either of two general situations: *either*  $x$  is self-evident to  $S$  *or* there is a set of propositions that fully grounds  $x$  for  $S$ . But we may consider the first of these a special case of the second, if we allow that the null set fully grounds self-evident propositions. Thus we may conclude that *a proposition  $x$  is evident to  $S$  if and only if there is a set of propositions that fully grounds  $x$  for  $S$ .*

If I am right thus far, the well-known triadic account of knowledge as correct, evident belief is then equivalent to an account (1) according to which  $S$  knows that  $p$  provided that both (a)  $S$  correctly believes that  $p$ , and (b) there is a set of propositions that fully grounds that  $p$  for  $S$ .

Supposing this account correct, every bit of knowledge would be supported by a pyramid such as the following.



Each node of such a pyramid is a proposition. Thus the apex node is the-proposition-that- $p_1$ , and the first terminal node from the left is the-proposition-that- $p_{111}$ .

Such "epistemic pyramids" must satisfy certain requirements:

1. (Let us call node  $x$  a "successor" (i.e., a direct successor) of node  $y$  relative to pyramid  $P$  provided there are  $A$  and  $B$  such that (i)  $A$  stands for  $x$  on  $P$  and  $B$  stands for  $y$  on  $P$ ; (ii)  $A$  and  $B$  are connected by a straight line on  $P$ ; (iii)  $B$  is closer to the apex node than  $A$ ; and (iv) there is no  $C$  such that  $C$  stands for a proposition on  $P$ , and  $A$ ,  $B$ , and  $C$  are connected by a straight line and  $B$  is closer to the apex node than  $C$ , and  $C$  is closer to the apex node than  $A$ .) The set of all nodes that (directly) succeed a given node must ground that node.
2. Each node must be a proposition that is evident to  $S$ .
3. If a node  $n$  is not a proposition that is self-evident to  $S$  then it must have successors, it must be succeeded by further nodes that ground the proposition  $n$ .
4. Each branch of an epistemic pyramid must terminate.

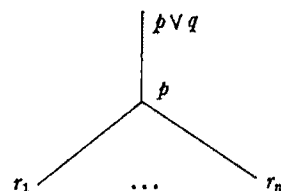
It follows that to each set that fully renders  $x$  evident to  $S$  there corresponds at least one epistemic pyramid for  $S$  and  $x$ . And we thus have a second account of knowledge that is also a close equivalent of the traditional account:

- (2)  $S$  knows that  $p$  iff
- (a) it is true that  $p$ ;
  - (b)  $S$  believes that  $p$ ; and
  - (c) there is an epistemic pyramid for  $S$  and the proposition that  $p$ .

However, it is a well-known fact that the tradi-

tional account of propositional knowledge is inadequate. Therefore, if the above account is really equivalent to the traditional account, it should be similarly inadequate. And so it is.

Consider the following pyramid (schema) for  $S$  and the proposition that  $p \vee q$ .



Supposing that  $S$  correctly believes that  $p \vee q$ , the three clauses in our account are satisfied. And yet  $S$  may not know that  $p \vee q$ , nevertheless. For it may be false that  $p$ , and the preceding diagram may be the *only* kind of epistemic pyramid available for  $S$  and the proposition that  $p \vee q$ , any other such pyramid being simply an "expansion" of the above, i.e., a pyramid obtainable by correctly adding further nodes. (Obviously, if there were another kind of pyramid, where instead of  $p$  we found  $q$ , then  $S$  might still know that  $p \vee q$ , in spite of the falsity of  $p$ .)<sup>7</sup>

A further restriction on pyramids immediately suggests itself: *to support knowledge epistemic pyramids must be non-defective, i.e., must contain only true nodes.*

According to Keith Lehrer the new requirement still leaves our account open to an interesting objection. Consider the following statements:

- $P_1$ : Mr. Nogot, who is in my office, owns a Ford;  
 $P_2$ : Mr. Havit, who is in my office, owns a Ford;  
 and  
 $H$ : Someone in my office owns a Ford.

And imagine "... that I have seen Mr. Nogot drive a Ford on many occasions and that I now see him drive away in it. Moreover, imagine that he leaves his wallet at my house and that I, being curious, examine its contents. Therein I discover a certificate asserting that Mr. Nogot owns the Ford I have just seen him drive away. This would supply me with evidence  $E$  consisting of true statements, which would completely justify my believing  $P_1$ , and, therefore,  $H$ . But now imagine that ...  $P_1$  is false (due to some legal technicality), and  $P_2$  is true, though I have no evidence for  $P_2$ . In this case, I do

<sup>7</sup> I have here translated the objection raised originally by Edmund L. Gettier against the traditional account of propositional knowledge. See his "Is Justified True Belief Knowledge?" *Analysis*, vol. 23 (1963), pp. 121-123 (reprinted in the Bobbs-Merrill reprint series).

not know  $H$ , but all of Mr. Sosa's requirements for knowledge might well be met."<sup>8</sup>

However, it is in general false that if one is justified in believing  $p$  and  $p$  entails  $q$ , then one is justified in believing  $q$ . I can surely believe the axioms of a theory and yet lack sufficient justification for believing many of the theorems. For many of the theorems may as yet be unproved.

Consequently, we must reject the general principle that if  $e$  provides good (inductive or deductive) grounds for  $p$ , and  $p$  entails  $q$ , then  $e$  provides good grounds for  $q$ . The point can also be put like this: even if your justifiably believing  $e$  and your believing  $p$  on the basis of your belief of  $e$  would together imply that your believing  $p$  is justified, and even if  $p$  entails  $q$ , it still would *not* follow that your justifiably believing  $e$  and your believing  $q$  on the basis of your belief of  $e$  would together imply that your believing  $q$  is justified. (Moreover, this would not follow even on the assumption that you do believe  $q$  and believe it on the basis of your belief of  $e$ .)

Thus I am puzzled by Lehrer's statement that  $E$  "... would completely justify my believing  $P_1$ , and, therefore,  $H$ ." (My emphasis.) I see no reason for accepting this inference. Would my believing  $E$  and my believing  $H$  on the basis of  $E$  imply that my believing  $H$  is justified? Not necessarily, for whether I am justified surely depends on whether I see a connection between  $E$  and  $H$ .<sup>9</sup> If I believe axioms of a theory and believe a theorem on their basis I still may not be justified in believing that theorem. For I may believe the theorem on the basis of the axioms mistakenly and without having seen a real connection.

Normally, we would establish a connection be-

tween  $E$  and  $H$  by way of the likes of  $P_1$  or  $P_2$ . Suppose that from among such statements only  $P_1$  is evident to you in some normal case. In that case you could make no connection between  $E$  and  $H$  except by way of a falsehood. Put another way, in that case there would be no non-defective pyramid for you and  $H$ . Therefore, our requirement would not be met by Lehrer's example after all. To be sure, there may be ways of connecting  $E$  and  $H$  without involving  $P_1$ . Perhaps there is some such way that does not violate the requirement, but I confess that I can't think of one.<sup>10</sup>

## II

Despite the merits of the present account, however, I believe it can be shown to be too narrow. For it allows as determinants of what one knows only what one correctly believes with rational justification. But reflection will reveal much else that is equally relevant.

In the first place, what one is rationally justified in believing obviously depends on the data in one's possession. But what data one has can depend on how much and how well one investigates. Consider, therefore, the following possibility. What if  $A$  is rationally justified in believing  $x$  given his body of data  $D_1$  whereas  $B$  is not rationally justified in believing  $x$  given his body of data  $D_2$ , where  $D_2$  includes  $D_1$  but is much more extensive as a result of  $A$ 's irresponsible negligence and  $B$ 's commendable thoroughness? The present account might unfortunately grant  $A$  knowledge while denying it to  $B$ , for  $A$ 's neglect so far has no bearing on any epistemic pyramid.

<sup>8</sup> Keith Lehrer, "Knowledge, Truth, and Evidence," reprinted from *Analysis*, vol. 25, in *Knowing*, ed. by M. D. Roth and L. Galis (New York, 1970), p. 171. The requirements to which Lehrer refers were made in my "Analysis of Knowledge that  $P$ ," *Analysis*, vol. 25 (1964-5), pp. 1-8 (reprinted in the Bobbs-Merrill reprint series). These requirements featured, in different terminology, the one now under review, viz., the requirement that for  $S$  to know that  $p$  there must be a nondefective pyramid of knowledge for  $S$  and the proposition that  $p$ . Indeed, the account of knowledge as correct belief buttressed by a nondefective pyramid is virtually my earlier account pruned of several inessential and confusing elements. I believe that the present account also enables us to deal with other difficulties reviewed in "Two Conceptions of Knowledge," *The Journal of Philosophy*, vol. 67 (1970), esp. pp. 63-66. The core idea involved has also been defended by Gilbert Harman, who argues that a condition for knowledge is that the lemmas be true, i.e., that one not reason *via* a falsehood. (See, e.g., "The Inference to the Best Explanation," *The Philosophical Review*, vol. 71 (1965), pp. 86-95; and "Knowledge, Reasons, and Causes," *The Journal of Philosophy*, vol. 6 (1970), pp. 841-856.) A recent defense of our core idea is "Knowledge without Paradox," by Robert G. Myers and Kenneth Stern (*The Journal of Philosophy*, vol. 70 (1973), pp. 147-161), which contains an interesting discussion of Lehrer's argument.

<sup>9</sup> For particularly strong requirements on the structure of such a connection, see James W. Lamb's "Knowledge and Justified Presumption," *The Journal of Philosophy*, vol. 69 (1972), pp. 123-127. This has been discussed by William Edward Morris in "Knowledge as Justified Presumption," *The Journal of Philosophy*, vol. 70 (1973), pp. 161-166.

<sup>10</sup> In favor of this reply to Lehrer's example, I would also urge that the proviso that one must not "reason" *via* a falsehood gives us the simplest, most natural solution to the Gettier problem, and that we should therefore let it fall only under the impact of a forceful counterexample. For a further defense of the proviso, see my forthcoming papers "On Our Knowledge of Matters of Fact" (*Mind*, 1974), and "Standard Conditions," the latter to appear in a volume in honor of Soren Hallden edited by the Philosophy Department of the University of Lund (Gleerup and Co., of Sweden, publishers).

We have considered a situation where someone lacks knowledge owing to his misuse of his cognitive equipment, either by letting it idle when it should be functioning or by busily employing it dysfunctionally. Another situation where someone lacks knowledge despite having rationally justified correct belief might be called the Magoo situation—where *S* lacks adequate equipment to begin with (relative to the question in hand: whether *p*).<sup>11</sup> It is because of this type of lack that despite his extensive experience with cable cars, Mr. Magoo does not know that his cable car will arrive safely when, unknown to him, bombs are raining all around it. Of course, even if you have less than 20–20 vision you can still know that there is an elephant in front of you when you see one there. So not just any defect will make your equipment inadequate for a judgment on the question whether *p*. I would venture that it must be a defect that prevents you from acquiring information that (i) a normal inquirer in the epistemic community would acquire in that situation and (ii) makes a difference to what you can reasonably conclude on the question whether *p* (or at least to how reasonably you can draw the conclusion).

The possibility of inadequate cognitive equipment requires a further and more striking departure from the traditional conception of knowledge. Earlier we considered a situation where, despite having warranted correct belief, someone lacks knowledge owing to his neglectful data-collection. There lack of knowledge could be traced back to epistemic irresponsibility, to substandard performance blamed on the investigator. In the present example, blame is out of place. By hypothesis, Magoo conducts impeccable “inquiry” both in arriving at his data and on the basis of his data. But he still falls short of knowledge, despite his warranted, correct belief. His shortcoming is substandard equipment, for which we may suppose him to be blameless. Hence something other than epistemic justification or correct belief can help determine what one knows or does not know. Even if one correctly believes that *p* with full rational justification and free of irrational or neglectful unbelief, one may still be in no position to know, because of faulty cognitive equipment.

In all of the foregoing cases, someone misses or is liable to miss available information which may be highly relevant and important and may make a difference to what he can conclude on the question in hand. In each case, moreover, he seems culpable or discredited in some sense: he would seem less reliable than otherwise for his role in any such case. But there appear to be situations where again someone misses available information with no culpability or discredit. Harman gives an example where *S* reads in a newspaper that some famous person has been assassinated, but does not read the next edition, where all reports of the assassination are denied by highly authoritative and trustworthy people. If practically the whole country reads the next edition and people don’t know what to believe, does *S* alone know of the assassination, provided the next edition is in fact a pack of lies?<sup>12</sup> I suppose we would be inclined to say that he does not know (especially if he had read the next edition, *he* would not have known what to believe). But what if only two or three people get a chance to read the next edition before it is recalled by the newspaper? Should we now say that out of the millions who read the first story and mourn the loved leader not one knows of his death? I suppose we would be inclined to say that the fake edition and the few deceived by it make no difference concerning what everybody else knows. It seems plausible to conclude that knowledge has a further “social aspect,” that it cannot depend on one’s missing or blinking what is generally known.

Our departures from the traditional conception of knowledge put in relief the relativity of knowledge to an epistemic community. This is brought out most prominently by the requirement that inquirers have at least *normal* cognitive equipment (e.g., normal perceptual apparatus, where that is relevant). But our new requirement—that inquirers not lack or blink generally known relevant information—also brings out the relativity. A vacationer in the woods may know that *p* well enough for an average vacationer, but he won’t have the kind of knowledge his guide has. A guide would scornfully deny that the tenderfoot really knows that *p*. Relative to the epistemic community of guides (for that area) the tenderfoot lacks rele-

<sup>11</sup> Thanks are due Jerome Shaffer for suggesting the problem posed by the Magoo situation—the situation of that unfortunate deaf and blind cartoon character who fortunately escapes disaster at every turn.

<sup>12</sup> Gilbert Harman, “Induction,” in Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief* (Synthese Library, 1970), esp. Sect. IV, pp. 95–97. Harman generously credits “The Analysis of ‘Knowledge that *p*,’” *op. cit.*, for this problem and for the “key to a solution.”



vant generally known information, and misses relevant data that the average guide would grasp in the circumstances.

(Understanding of these departures from the traditional account may perhaps be enhanced if we reflect that the honorific term "knowledgeable" is to be applied only to those who are reliable sources of information, surely an important category for a language-using, social species.)

We have now taken note of two types of situation where correct, fully warranted belief falls short of knowledge owing to no neglect or faulty reasoning or false belief. Despite commendable thoroughness and impeccable reasoning unspoiled by falsehood, one may still fail to be "in a position to know," owing either to faulty cognitive equipment or to missed generally known information. I am not suggesting that these are the only ways to be out of position to know. I have no complete list of epistemic principles describing ways of arriving at a position to know or of being blocked from such a position. My suggestion is only that there are such principles, and that in any case we must go beyond the traditional emphasis by epistemologists on warrant and reasoning as determinants of knowledge. Despite the importance of warranted correct belief in determining what we know, the Gettier examples show that it is not alone enough to guarantee knowledge. What is more, warranted correct belief supported by reasoning *unspoiled by falsehood* seems immune to Gettier examples, but it may still fall short of knowledge. We have considered several types of such shortfalling.

My conclusion is that to understand knowledge we must enrich our traditional repertoire of epistemic concepts with the notion of *being in a position to know* (from the point of view of a *K*, e.g., a human being). Thus a proposition is evident (from the point of view of a *K*) to a subject only if *both* he is rationally justified in believing it *and* he is in a position to know (from the *K* point of view) whether it is true. It may *be* (and not just appear) evident to Magoo from *his* point of view that he will reach the other side safely, but it seems wrong to say of Magoo as he steps into the cable car with bombs raining all around that it is quite evident to him that he will arrive safely. It seems wrong *for whom* to say this? For one of us, naturally; that is, for a normal human from *his* point of view. And since a normal human could not help seeing and hearing the bombs, from the human point of view Magoo is not in a position to know that he will arrive safely, inasmuch as he is missing relevant information that

a normal human would gather in the circumstances. Hence Magoo does not have *human* knowledge that he will arrive safely, for it is not evident to him from the human point of view that he will so arrive.

Our latest account was this:

- (3) *S* knows that *p* iff
  - (a) it is true that *p*;
  - (b) *S* believes that *p*; and
  - (c) there is a non-defective epistemic pyramid for *S* and the proposition that *p*.

It will be recalled that every node of such a pyramid must be true *and* evident. And for every node *n* that has successors, the successors must serve as grounds that give the subject *S* rational warrant for believing *n*. What now seems too narrow about this account emerges with the explanation of what a pyramid of knowledge is, and of what the evident is. For in this explanation what is evident to *S* is identified with what *S* is rationally justified in believing. But it now seems plain that for *x* to be evident to *S*, *two* conditions must be satisfied: (i) that *S* be rationally justified in believing *x*, and (ii) that *S* be in a position to know whether *x* is true. And we must also take note of the relativity of knowledge to an epistemic community. Let us therefore replace (3) with the following:

- (4) *S* knows (from the *K* point of view) that *p* iff
  - (a) it is true that *p*;
  - (b) *S* believes that *p*; and
  - (c) there is a non-defective epistemic pyramid (from the *K* point of view) for *S* and the proposition that *p*.

Every node of such a pyramid must now be true and evident from the *K* point of view.

Normally when epistemologists discuss knowledge (of the colors and shapes of surrounding objects, of one's own or one's neighbor's mental states, and so on), they plainly do so from the *human* point of view. But other points of view are possible even in ordinary conversation. The expert/layman distinction is replicable in many different contexts, and with each replication we have a new epistemically relevant distinction in points of view, with expert knowledge on one side and layman knowledge on the other.

Neither Magoo nor the newspaper reader who alone has not seen the new edition is in a position to know (from the human point of view) about the relevant subject matter. Thus we can understand their ignorance and, by parity of reasoning, the

ignorance of all those who are out of position to know that  $p$  because they lack either adequate cognitive equipment or relevant information that is generally known to those who have taken an epistemic stand on the question whether  $p$  (where to suspend judgment is to take an epistemic stand, whereas to be totally oblivious to the matter is not).

But the new account, with its broader conception of the evident, seems blatantly circular. For  $x$  must be evident to you if you are to know that  $x$  is true. But now  $x$  is evident to you only if you are in a position to know whether  $x$  is true. And what is it to be in a position to know except to be in a position such that if you believe correctly then you do know? Is there any defense against this charge of circularity? Let us begin by conceding—what is in fact questionable—that you are in a position to know if and only if you are in a position such that you have only to believe correctly in order to know. All the same, this need not be offered as an analysis but only as a truth. If that is not the analysis of being in a position to know, however, what is? Our account escapes the noose of circularity only to face a pit of obscurity.

A defense can be given in two stages. In the first place, if our account seems obscure at first sight, the traditional account seems equally obscure. What is the “rational justification” of which it speaks? This is particularly puzzling because it must be understood as a *sui generis* type of intellectual or theoretical justification, so as to rule out whatever practical considerations may militate for or against belief. In the second place, insofar as rational justification may be clarified it surely must be in terms of the principles of reasoning or epistemic principles that define its area of application. And the same may be done for being-in-a-position-to-know. Hence insofar as our account seems vague or obscure, the traditional account seems equally so, and insofar as the traditional account allows of clarification our account does so as well.

Thus  $S$  is in a position to know that  $p$  if it is self-evident to  $S$  that  $p$ , . . . , and so on, drawing the content of our explanation from the relevant epistemic principles. Two such principles have been suggested as solutions to as many problems: (a) that to be in a position to have human knowledge whether  $p$  you must not miss any crucial information that a normal human in your situation would gather; and (b) that to be in a position to have human knowledge whether  $p$  you must not miss any crucial

information that is generally known to those who have taken an epistemic stand on the question whether  $p$ . (Information is crucial relative to your knowing whether  $p$  provided that adding that information to your evidence base would induce a fall in the epistemic status of your belief either that  $p$  or that not- $p$  to such an extent that where previously that belief was rationally justified it no longer is so.) It is such principles that give content to the idea of being in a position to know.

Doubtless we cannot hope here for such precision as is attainable in explaining, e.g., what it is to be a chess bishop. But chess is defined by a set of artificial conventions, whereas it is doubtful that epistemic principles are either artificial or conventional. At present, in any case, our understanding of what it is to be rationally justified or of what it is to be in a position to know is about as vague as our understanding of what it is to be moral, or well-mannered, or to speak grammatically. It would be very difficult to remove such vagueness, and it might turn out to be impossible.<sup>13</sup> In that case, Russell's view that knowledge is an inherently vague notion would have been vindicated. Meanwhile it is surely wise to elicit as much precise content as possible from our epistemic practice, for content unelicited is missed understanding.

### III

What it is for  $S$ 's belief that  $p$  to be fully grounded had been explained earlier by means of our epistemic pyramids. That answer points in the right direction, but it needs to be made more precise. (Perhaps the main requirement is clarifying the grounding relation.) Moreover, we have found that a fully grounded correct belief is not necessarily knowledge, and this for at least two reasons: (i) it may rest directly or indirectly on some false ground, and (ii) the believer may not be in a position to know.

We have tried to allow for these possibilities by broadening epistemic pyramids, by making room for our new epistemic notion of being-in-a-position-to-know, and by noting that to support knowledge epistemic pyramids must be non-defective, i.e., must contain no false nodes. But pyramids are objectionable for other reasons as well: (i) they may mislead by suggesting that terminal nodes provide a “foundation” in one or another undesirable sense, or by suggesting that terminal nodes must

<sup>13</sup> Cf. Malcolm, *op. cit.*, pp. 114–115.

come first in time, so that one may later build on them; (ii) more seriously, there is an unacceptable vagueness in the very idea of such a pyramid, which derives mainly from the vagueness of the "grounding" relation in terms of which pyramids were defined. What follows is an attempt to solve these problems by switching pyramids upside down into trees, and by providing a more precise definition of such trees. To achieve greater precision, however, we must first enrich our conceptual resources. Each of the three main concepts to be defined—(I), (II), and (III)—seems to me a source of light, and by this light we shall then trim our trees to a more precise outline.

- (I)  $\alpha$  fully validates  $S$ -epistemic proposition  $e$  if and only if there is a set of subsets of  $\alpha$  that form a sequence,  $C_1, \dots, C_n$ , such that (i)  $C_1$  validates  $e$ , (ii) each  $S$ -epistemic member of any subset  $C_i$  is validated by the immediate successor of  $C_i$ ,  $C_{i+1}$ , and (iii) no member of the last subset,  $C_n$ , is  $S$ -epistemic.

This account of full validation makes use of a battery of technical concepts. Among them are *validation* and *S-epistemic proposition*, each of which requires explanation.

- (II) *S-epistemic propositions*: of these (i) some are to the effect that some other proposition  $x$  has some epistemic status relative to  $S$  (i.e., to the effect that  $S$  believes  $x$  and how reasonably); and (ii) others are to the effect that  $S$  is in a position to know whether a certain proposition is true. (i) and (ii) are the varieties of "positive  $S$ -epistemic propositions." Logical compounds of such propositions are also  $S$ -epistemic, but no other propositions are  $S$ -epistemic.
- (III) *Epistemic validation*. Let  $A$ ,  $B$ , and  $C$  be sets of propositions.
1.  $A$  epistemically implies that  $p$  if and only if  $A$  and the epistemic principles together logically imply that  $p$ , but neither does so alone.

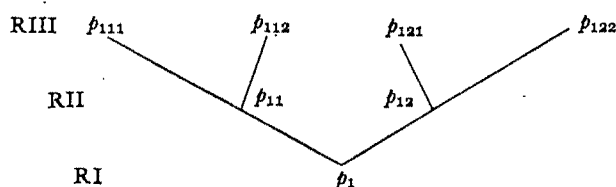
2.  $A$  absolutely validates  $S$ -epistemic proposition  $e$  if and only if  $A$  has only true members and epistemically implies  $e$ .
3.  $A$  confirms  $S$ -epistemic proposition  $e$  if and only if (a)  $A$  is a set with only true members, and (b) supposing the true, positive  $S$ -epistemic propositions are at most those entailed or absolutely validated by  $A$ ,  $e$  follows epistemically from this together with  $A$ .
4.  $A$  validates  $S$ -epistemic proposition  $e$  if and only if  $A$  confirms  $e$ ,  $A$  has only true members, and for every  $B$  with none but true members, such that  $A \cup B$  does not confirm  $e$ , there is a  $C$  with none but true members such that  $A \cup B \cup C$  does confirm  $e$ .<sup>14</sup>

(Note that 3 is included only for convenience, since  $A$  confirms  $p$  if and only if  $A \cup \{\text{that the true, positive } S\text{-epistemic propositions are at most those entailed or absolutely validated by } A\}$  absolutely validates  $p$ .)

That completes the explanation of full validation (defined above). Note that what is fully validated by a set is always an  $S$ -epistemic proposition; and that the propositions in the validating set need not all be  $S$ -epistemic and that indeed perhaps none is  $S$ -epistemic. For example, some may be "self-certifying" or "self-discrediting" by specifying, with respect to a proposition  $x$ , that  $S$  correctly believes  $x$ , where this is sufficient for  $x$  to have a certain (positive or negative) epistemic status for  $S$ , given the epistemic principles; other *non-epistemic* members of a validating set may specify  $S$ 's perceptual or cognitive basis for a belief he has; others may certify the adequacy of  $S$ 's cognitive equipment relative to the epistemic community; and others yet may guarantee that  $S$  is not missing relevant generally known information.

(For convenience, in what follows let us terminologically stipulate that  $\alpha$  (fully) validates that a proposition is evident, certain, reasonable, etc., for a subject when and only when the set  $\alpha$  "(fully) renders certain, reasonable, etc.," that proposition for that subject; and that  $\alpha$  (fully) validates that  $S$

<sup>14</sup> The notion of epistemic validation makes the present proposal a "defeasibility analysis of knowledge," in Marshall Swain's terminology. Swain finds the earliest such use of defeasibility in "The Analysis of 'Knowledge that p'," *op. cit.* [see M. Swain, "Knowledge, Causality, and Justification," *The Journal of Philosophy* (1972), n. 2]; and he traces its use through a dozen papers on knowledge in his forthcoming "Epistemic Defeasibility," *American Philosophical Quarterly*, vol. 11 (1974). Most recent developments include Swain's "Alternative Analysis of Knowledge," forthcoming in *Synthese*, and "Epistemic Defeasibility." Finally, I have found R. M. Chisholm's "On the Nature of Empirical Evidence" especially helpful in thinking about epistemic validation. It appeared originally in *Experience and Theory*, ed. by Lawrence Foster and J. W. Swanson (Amherst, 1970); and a revised version appears in *Empirical Knowledge* by Chisholm and R. J. Swartz (New York, 1970).



is in a position to know that  $p$  when and only when  $\alpha$  “(fully) puts  $S$  in a position to know that  $p$ .”)

There is some simple reasoning behind the definition of full validation and behind the idea that if  $S$  knows that  $p$  then there must be a set that fully validates that it is evident to  $A$  that  $p$ . The reasoning boils down to this: Supposing that both circular justification and an infinite regress are ruled out, it would seem that a chain of justification must eventually reach a last link that either is self-validating or helps validate further links without itself requiring validation.

Let us emphasize, however, that this will not commit one to a picture of knowledge according to which there is a bedrock of self-evident propositions. It is perfectly consistent with the present theory that part of what makes *any* proposition evident be its coherence with a network of mutually supporting propositions. Since there is bound to be a multitude of such coherent networks, however, a non-arbitrary narrowing of the field must be supported by something other than coherence. And this is where the last subset in the sequence of subsets of a fully validating set may be invoked. However, it is not required that the last subset include any self-evident belief; for example, it may contain some merely self-certifying correct belief, i.e., a proposition specifying that some other proposition  $x$  is correctly believed, where this suffices for  $x$  to have *some* positive epistemic status, however lowly: acceptability, for instance (i.e., being such that not believing it is *not* preferable to believing it).

The concepts recently defined enable us to construct an account (5) according to which  $S$  knows that  $p$  provided that both (a)  $S$  correctly believes that  $p$ ,<sup>15</sup> and (b) there is a set of propositions that

fully and nondefectively renders it evident to  $S$  that  $p$  (where a set “nondefectively renders it evident to  $S$  that  $p$  if and only if it does so without attributing to  $S$  any false belief”).<sup>16</sup>

Supposing this account correct, every bit of knowledge has a tree such as the following, the “ranks” of which (RI, RII, and RIII) correspond to our epistemic subsets discussed above (in the definition of full validation).

Note that each node of such a tree is a proposition. Thus the “root” node is the-proposition-that- $p_1$ , and the first terminal node (from the left) the-proposition-that- $p_{111}$ .<sup>17</sup>

There is an important difference between these trees and our earlier pyramids. Except for terminal nodes, every node of a tree is an  $S$ -epistemic proposition, whereas not a single node of a pyramid need be epistemic at all. Pyramids display propositions that are evident to  $A$  (*not* propositions that such and such other propositions are evident to  $S$ ), and they also show which propositions ground (for  $S$ ) any proposition for which  $S$  has grounds. Trees display true epistemic propositions concerning  $S$  and they also show what “makes these propositions true” *via* epistemic principles (i.e., what successors “validate” these propositions). Trees do this for *every* epistemic proposition that constitutes one of its nodes. That is to say, trees contain no epistemic terminal nodes. It is in this sense that trees provide *complete* epistemic explanations of the truth of their root nodes.

The following requirements flow from the fact that trees are intended to correspond to the fully validating sets defined by (I) above, and from the fact that the ranks of trees are to correspond to the subsets  $C_1, \dots, C_n$  of (I).

<sup>15</sup> Whether knowledge entails belief at all is of course a vexed question of long standing, but there is no room for it here. A helpful and interesting discussion is found in Keith Lehrer’s “Belief and Knowledge,” *The Philosophical Review*, vol. 77 (1968), pp. 491–499.

<sup>16</sup> Here, in the account of concepts (I)–(III) above, and in what follows, the relativity of knowledge to an epistemic community is left implicit, as it normally is in ordinary thought and speech.

<sup>17</sup> Strictly speaking, what we have here is obviously a *partial tree schema*. For convenience, however, I speak of trees even when I mean partial tree schemata. Also, it should not be thought that every tree must have exactly three ranks (RI, RII, and RIII). On the contrary, a tree may have any number of ranks, so long as it has more than one.

1. (Let us call node  $x$  a "successor" (i.e., a *direct* successor) of node  $y$  relative to tree  $T$  provided there are  $A$  and  $B$  such that (i)  $A$  stands for  $x$  on  $T$  and  $B$  stands for  $y$  on  $T$ ; (ii)  $A$  and  $B$  are connected by a straight branch on  $T$ ; (iii)  $B$  is closer to the root node than  $A$ ; and (iv) there is no  $C$  such that  $C$  stands for a proposition on  $T$ , and  $A$ ,  $B$ , and  $C$  are connected by a straight branch and  $B$  is closer to the root node than  $C$ , and  $C$  is closer to the root node than  $A$ .) The set of all nodes that (directly) succeed a given node must validate that node.
2. If a node  $n$  is an  $S$ -epistemic proposition, then  $n$  must have successors, it must branch out into a set of further nodes that validate  $n$ . (Trees are supposed to correspond to sets that fully validate  $S$ -epistemic propositions, and any such set is to provide, in a sense, a "complete" explanation of how the  $S$ -epistemic proposition comes to be true.)
3. The proposition that  $p_1$  must be  $S$ -epistemic.
4. Each branch of a complete tree of evidence must terminate. (And in some cases the terminal node specifies that  $S$  correctly believes  $x$ , and is the sole successor of its immediate predecessor, which is to the effect that  $x$  has a certain epistemic status for  $S$ .)
5. A non-defective tree of evidence for  $S$  must attribute no false belief to  $S$ . This applies not only to terminal nodes but also to non-terminal nodes. We know that a non-terminal node must be an  $S$ -epistemic proposition, given the nature of the consequents of epistemic principles and the fact

that it is  $S$ 's knowledge that is in question.<sup>18</sup> That is to say, each non-terminal node must be a proposition concerning a certain other proposition (or certain other propositions) and either (i) to the effect that it is (they are) believed by  $S$  and how reasonably, or (ii) to the effect that  $S$  is in a position to know whether it is (they are) true, or (iii) some logical compound of propositions of types (i) or (ii). Propositions of types (i) or (iii) may attribute belief to  $S$ . If the tree is to be non-defective, no such belief can be false.

Finally, if the root node of a non-defective tree of evidence is the proposition that it is evident to  $S$  that  $p$ , let us call such a tree not just a tree of evidence, but a *tree of knowledge* for  $A$  and the proposition that  $p$ .

In conclusion, we may now be sure that to each set that fully renders it evident to  $S$  that  $p$  there corresponds at least one tree of knowledge for  $S$  and the proposition that  $p$ . And we thus have an equivalent for our account of knowledge (5) as correct belief in what has been fully and non-defectively rendered evident:

$S$  knows that  $p$  iff

- (a)  $S$  correctly believes that  $p$ ; and
- (b) there is a tree of knowledge for  $S$  and the proposition that  $p$ .<sup>19</sup>

*Brown University*

*January 25, 1973*

<sup>18</sup> But it should be noted that terminal nodes are not restricted to the last (or top) rank of a tree. On the contrary, any number of them may appear in any rank.

<sup>19</sup> I am indebted to Roderick M. Chisholm, Jaegwon Kim, Alvin I. Goldman, Robert Merrihew Adams, Stephen Leeds, and Robert J. Swartz. Work on this paper was supported in part by NSF Grant GS-2953.

## V. THEORIES OF INTRINSIC VALUE

WARREN S. QUINN

IT can be said without exaggeration that there is no major ethical theory of any sort into which the concepts of intrinsic value or valuation do not enter in some way or other. Teleologists typically base their conception of right action on supposed pre-moral intrinsic values, while deontologists often ground theirs in the alleged intrinsic moral value (fittingness, nobility, etc.) of certain familiar act types. Even those who doubt the reality of objective intrinsic values cannot so easily dispense with intrinsic *valuation*—i.e., that a person may like and prefer things for their own sake. Such notions enter essentially into the concepts of rationality and prudence, since the prudent or rational choice is by common definition the one likely to bring about conditions which its agent intrinsically prefers. This paper will, however, ignore the admittedly interesting distinction between value and valuation since the theories to be discussed apply equally well (*mutatis mutandis*) to the analysis of either set of concepts.

These theories are squarely metaethical, and while they may have some implications which are normative, they aim at conceptual illumination rather than personal guidance. Despite their obvious differences each regards the potential objects of intrinsic value or valuation to be propositional. In addition, several of the better known analyses involve the idea of states of affairs or propositions (I shall use these terms interchangeably) in relation to possible worlds. In dealing with such analyses I shall give the notion of a possible world a naive, prelinguistic interpretation. And whenever it is relevant to the theory being considered, a proposition will be identified with a set of such possible worlds, viz., the set of all and only the worlds in which the proposition is true.

This identification gives rise to a logical scheme which may be unfamiliar but which is easily assimilated. The negation of a proposition  $p$  is to be identified with the set of all the worlds excluded from  $p$ . A disjunction  $p \vee q$  is to be thought of as the union of the set identified with  $p$  and the set identified with  $q$ ; while the conjunction  $p \& q$  is taken as their intersection. (Note that any subset of  $p$  will

be called a "disjunct" or "disjunctive part" of  $p$ , and if nonequivalent to  $p$  it will be called a "narrower" disjunct of  $p$ .) Similarly,  $p$  will be taken to entail  $q$  if and only if  $p$ 's set of worlds is contained within (is a subset of)  $q$ 's, and two propositions will be regarded as logically equivalent just in case they have the same member worlds. Note also that a proposition may contain any number of possible worlds: propositions which contain only one world—to be called "unit propositions"—function importantly in two of the theories to be discussed. Finally, two propositions will be called "logically compatible" when they intersect in some possible world, "logically independent" when they intersect but neither is wholly contained within the other, and "logically disjoint" or "logically incompatible" when they do not intersect at all.

Criticism will sometimes take advantage of the fact that each of the metaethical theories to be considered has incidental normative implications which, if sufficiently implausible, provide a means of calling the theory into question. The normative touchstones to be relied upon in such criticism are first, that it is intrinsically good that a man be happy, at least so long as his happiness is non-malevolent—i.e., does not result from the thought of someone else's unhappiness; second, that it is intrinsically better that a man be more rather than less nonmalevolently happy; and third, that the state of affairs in which two different persons  $A$  and  $B$  are both nonmalevolently happy is intrinsically better than the state in which  $A$  is non-malevolently happy. (For convenience, I shall henceforth use "happy" as short for nonmalevolently happy.)

In both illustration and criticism I shall adopt the expedient of supposing that happiness and unhappiness can be measured. These imaginary measurements will be indicated by attaching numerical subscripts to the appropriate predicates, e.g., "happy<sub>2</sub>," "happy<sub>-3</sub>" (a negative subscript indicating a negative degree of happiness or, in other words, a positive degree of unhappiness), and "happy<sub>0</sub>" (neither happy nor unhappy). To further facilitate illustration, I shall restrict

my subjects to two imaginary persons, Jane and Bill, representing Jane's being happy or unhappy to degree  $n$  with the convenient symbols ' $j_n$ ' or ' $j_{-n}$ ' and Bill's being happy or unhappy to degree  $n$  with the symbols ' $b_n$ ' or ' $b_{-n}$ '. For example, that Jane is happy<sub>2</sub> will be represented ' $j_2$ ' and that Bill is unhappy<sub>5</sub> as ' $b_{-5}$ '.

The four theories or partial theories to be considered in Sect. I represent clearly distinguishable analyses. The first three build on the work of others but are not their responsibility. The processing, expansion, and relocation of these borrowed ideas will be apparent to readers familiar with the originals. The fourth theory, however, is taken directly from Gilbert Harman. As to content, the first two theories give principles by which the intrinsic value of wider propositions can be determined from the values of their narrower disjuncts. By a stretch of terminology these value characterizations could be called "recursive." The next two theories, on the other hand, attempt to reduce intrinsic value to still more basic value concepts, and thus might be called "reductionist." Sect. II will present my own solution to the analytical problems raised by these four theories—a solution which borrows eclectically from both the "recursive" and "reductionist" viewpoints.

## I

(1) The first theory to be considered is based on the work of von Wright.<sup>1</sup> It differs, however, from the original in several significant ways. Von Wright's account is relativized to persons and times, is restricted to states fully describable by occasion sentences, and involves complications concerning the change from one state of affairs to another. My version is both simplified and expanded but clearly retains the most central idea—viz., that  $p$  is intrinsically better than  $q$  if and only if any pair containing a world from  $p \& -q$  and a world from  $q \& -p$  which two worlds are alike in all other respects, is such that the subset containing only the  $p \& -q$  world is intrinsically better than the subset containing only the  $q \& -p$  world. (For shorthand, I shall drop the talk of singleton sets of worlds—i.e., unit propositions—and simply speak of the worlds themselves; moreover when dealing with a logically incompatible  $p$  and  $q$  I shall simply speak of the  $p$ -world and the  $q$ -world, rather than the

$p \& -q$  world and the  $q \& -p$  world.) Given value orderings for every pair of possible worlds, the above formula enables one to determine the relative intrinsic merits of any two propositions. Moreover, a definition of intrinsic goodness is generated, according to which a proposition is intrinsically good if and only if it is intrinsically better than its negation.

The tricky problem in penetrating this account of betterness is to interpret the italicized "alike-ness" condition in a plausible way. For a pair of worlds one member of which belongs to  $p$  and the other to  $q$  to be alike in all other respects would seem at first to mean that they must be materially equivalent with respect to every other proposition, i.e., every proposition not logically equivalent to either  $p$  or  $q$ . But this interpretation of the likeness condition leads to impossible demands. For example, in comparing  $j_1$  and  $j_2$  we would be forced to make each relevant ( $j_1:j_2$ ) pair of worlds<sup>2</sup> materially equivalent with respect to such a proposition as  $j_1 \vee j_0$ . But no pair of ( $j_1:j_2$ ) worlds can be equivalent with respect to this proposition, and hence the basis for the comparison disappears. It therefore seems necessary to interpret the likeness condition as having within its scope only propositions which logically speaking *can* be held materially equivalent for at least some pair of ( $p \& -q:q \& -p$ ) worlds.

Unfortunately, this interpretation will not yield any comparisons either—at least for ordinary propositions which contain more than one possible world apiece. For any putatively matched pair of ( $p:q$ ) worlds from two such ordinary propositions  $p$  and  $q$  (for convenience, assume them to be logically incompatible) there is always a distinct proposition  $r$  which is equivalent in some ( $p:q$ ) pairs but is not equivalent in the pair in question. Let  $w_p$  and  $w_q$  be any two worlds from  $p$  and from  $q$  respectively. And let the proposition that world  $w_p$  is actual be  $p_w$  and the proposition that world  $w_q$  is actual be  $q_w$ . It is clear that the disjunction  $p_w \vee (q \& -q_w)$  is compatible with both  $p$  and  $q$ —i.e., is true in some ( $p:q$ ) pairs. But it is also a proposition which holds of  $w_p$  but not of  $w_q$ . Since this construction is perfectly general for propositions containing multiple worlds, it follows that for any such pair of ( $p:q$ ) worlds there will always be a distinct proposition  $r$  which cannot be held equal.

It may be that this impasse can be breached by

<sup>1</sup> G. H. von Wright, *The Logic of Preference* (Edinburgh, 1963), esp. p. 34.

<sup>2</sup> The ' $(\phi:\psi)$ ' notation is used to indicate an ordered pair of worlds whose first member is from proposition  $\phi$  and whose second member is from proposition  $\psi$ .

further restrictions on the class of propositions to be held equal, but every diminution of this class brings a corresponding diminution in the security of the comparisons founded upon it. But even more to the point, I cannot see what such further restrictions could plausibly consist in. For this reason it might seem that the theory ought to temper its demand for perfectly matched pairs (with respect to some antecedently specifiable class of propositions) and should look instead to *maximally matched pairs*—pairs equivalent with respect to as many distinct propositions as possible. The theory thus revised would hold that  $p$  is intrinsically better than  $q$  if any maximally similar ( $p \& -q : q \& -p$ ) pair has a better first than second member, while the definition of intrinsic goodness would presumably be unaffected. Any attempt to develop this version of the theory would of course be burdened by the very unclear notion of maximal similarity. But no matter how this notion is explicated in detail, the following problems would seem to remain.

First, it is unlikely that any such theory can show that, e.g.,  $j_2$  is intrinsically better than its negation  $-j_2$ . In the search for ( $j_2 : -j_2$ ) pairs which are materially equivalent with respect to as many propositions as possible, what are we to say of pairs whose  $-j_2$  member is in, e.g.,  $j_{10}$ ? It would not be possible automatically to reject such a pair as not maximally similar since every ( $j_2 : -j_2$ ) pair will differ by some such  $j_n$ . That is, every  $-j_2$  world in which Jane exists will have some proposition of the form  $j_n$  (where  $n \neq 2$ ) true in it which proposition cannot be true in any  $j_2$  world. Some of these "unequal" propositions, such as  $j_{10}$ , are better than  $j_2$  while others, such as  $j_{-10}$ , are worse. The salient point is that the presence of such a disparity between a pair of ( $j_2 : -j_2$ ) worlds could not show that the worlds fail to have as many propositions in common as possible. It therefore seems inevitable that some of our maximally similar ( $j_2 : -j_2$ ) pairs will have their  $-j_2$  member in  $j_{10}$  and that some of these latter pairs will therefore have a better  $-j_2$  member. This means that  $j_2$  is not intrinsically better than its negation and also, given the theory's definition of intrinsic goodness, that  $j_2$  is not intrinsically good.

A second criticism stems from the fundamental assumption that in comparing the intrinsic values of the distinct propositions  $p$  and  $q$  one is really comparing the value of  $p \& -q$  with that of

$q \& -p$ . This has some plausibility in the event that  $p$  and  $q$  are logically independent (e.g.,  $j_1$  and  $b_1$ ) and even more in the event that  $p$  and  $q$  are logically disjoint (e.g.,  $j_1$  and  $j_2$ ). However, the assumption must be questioned in the case in which  $p$  and  $q$  are neither logically independent nor disjoint, i.e., where one proposition entails the other. Suppose, for example, that  $q$  entails  $p$  but that  $p$  does not entail  $q$ ; then to ask the apparently sensible question whether  $p$  is better or worse than  $q$  is, on the above assumption, to ask the apparently senseless question whether  $p \& -q$  is better or worse than the contradiction  $q \& -p$ .

One way around this difficulty would be to stipulate that in such cases a comparison of  $p$  with  $q$  is tantamount to a comparison of  $p \& -q$  with  $q$ . In other words, to ask whether the whole is better or worse than the part is to ask whether the remainder of the whole is better or worse than the part. However, this proposal seems a priori objectionable on the ground that the part of  $p$  we are permitted to compare with  $q$  does not represent, so to speak, a fair sample of  $p$ . Suppose, for example, we compare the disvalue of Jones hurting Smith with that of Jones hurting Smith but not enjoying it. On the present assumption this amounts to a comparison of Jones hurting Smith *and* enjoying it with Jones hurting Smith but not enjoying it. Yet surely it is possible—indeed plausible—to deny that Jones hurting Smith and enjoying it is for purposes of comparison evaluatively equivalent to Jones hurting Smith (sans qualification).

The alternative of dealing with these nested propositions by admitting comparisons of worlds from any parts of  $p$  and  $q$  is even worse. In attempting to keep as many propositions as possible equal between our ( $p : q$ ) pairs we would then have to keep  $q$  equal, which means in effect that we would be comparing  $p$  worlds which are  $q$  with  $q$  worlds which are  $p$ , and it is hard to see what such comparisons could reveal about the respective merits of  $p$  and  $q$ . In summary, therefore, the "maximal similarity" modification of the present theory would seem to yield strikingly counterintuitive results on the one hand and to give no satisfactory way of comparing nested propositions on the other.

(2) The second view is modeled on Rescher's theory of "first order" goodness and betterness.<sup>3</sup> While the original is not offered as a specific account of *intrinsic* value, the elegance of the model

<sup>3</sup> Nicholas Rescher, "Semantic Foundations for the Logic of Preference" in Nicholas Rescher (ed.), *The Logic of Decision and Action* (Pittsburgh, 1966).



invites application. When transformed into our semantic framework, it would require the assignment of an intrinsic value to every unit proposition and would calculate the intrinsic value of every other proposition by averaging the values of the unit propositions it contains. However, the propositions of greatest interest (e.g.,  $j_1$ ) contain an infinite number of possible worlds and an infinite number of values cannot be averaged. Thus, instead of making the basic assignments to unit propositions we should perhaps look to sets of possible worlds whose members are from our vantage point indistinguishable. Any set which has this character may be called a "cell." Such cells are most naturally individuated by means of the semantically consistent state descriptions of our language. This method guarantees that they will be finite in number and hence averageable. If an intrinsic value is assigned to each cell then the intrinsic value of any proposition expressible in the language as a disjunction of state descriptions may be calculated as the average of the values of its member cells.

Despite its appealing simplicity, the present view is subject to certain doubts. For one thing, the type of value in question imposes certain constraints on the initial assignment of value to the cells. This is because, apart from metaphysical optimism or pessimism, it is difficult to conceive of the necessary proposition  $N$  as being anything other than intrinsically neutral. Therefore the sum of the values of all the cells must be zero, and from this it is easy to see that any partition of  $N$  into a proposition and its negation will be such that unless both are indifferent the intrinsic value of one will be the negative of the intrinsic value of the other. Thus from the fact that  $j_2$  is good,  $-j_2$  must be bad; and from the fact that  $j_{-2}$  is bad,  $-j_{-2}$  must be good. But both  $-j_2$  and  $-j_{-2}$ , in that neither entails Jane's happiness or unhappiness, are commonly regarded as neutral.

There is a further doubt as well. The theory asks us to reckon the value of, e.g.,  $j_1$  in terms of the values of the unit propositions or cells it contains. In other words, it treats these latter values as in some way prior to the value of  $j_1$ . But surely from an epistemological point of view the priority seems the reverse. We only know the value to be placed on a world in  $j_1$  and hence on its unit set because we know the values of the states of affairs (such as

$j_1$ ) which are true in it. Moreover, I would argue that our conception of intrinsic goodness is such as to include the possibility that the value of a narrower proposition is to be understood in terms of the values of the wider ones which it entails (and that this possibility is certainly realized in the relation between a unit set in  $j_1$  and  $j_1$  itself). The present theory, in requiring us to conceive our values from inside out, contradicts the strong appearance that they are just as frequently conceived in the reverse direction.

(3) The first two theories treat propositional intrinsic value as conceptually primitive, i.e., not to be explained in terms of other types of value or in terms of objects other than propositions. In this they contrast with the two theories which remain to be discussed in Sect. I. The first of these, based on mere hints in Chisholm and Sosa, uses the analogy between the relation of a proposition to a member world and the relation of a property to a subject possessing it.<sup>4</sup> For example, the proposition that Jane is happy may be likened to an attribute which holds of those worlds in which the proposition is true and which does not hold of those worlds in which the proposition is false. A proposition is intrinsically good, according to this theory, if and only if it could be regarded as a *good-making property of worlds*. Chisholm and Sosa's remarks do not further develop this interesting idea; the following projection is therefore my own responsibility.

On consideration, the theory suggested would reduce the notion of intrinsic goodness to two prior value notions: that of the goodness of worlds and that of good-makingness. The goodness of worlds itself appears to be a type of innate value but is distinguished from intrinsic value by the fact that it holds of particular worlds rather than propositions. For present purposes its other characteristics may be left unspecified. The other value notion involved in the analysis, viz., good-makingness, is found in other parts of value theory. On the usual analysis, a property is good-making first with respect to particulars of a specific kind and second with respect to a specific type of value. A familiar example would be honesty as good-making with respect to *persons* and *moral worth*. It is important that good-makingness not be confused with a mere causal or evidential relation. The good-making property is *constitutive* of whatever value belongs to the objects which possess it

<sup>4</sup> Roderick Chisholm and Ernest Sosa, "On the Logic of 'Intrinsically Better'," *American Philosophical Quarterly*, vol. 3 (1966), pp. 244-249.

and hence constitutive of the goodness of the good objects which possess it. Being born of good parents may be a cause of or evidence for a person's moral worth—but it cannot constitute it. In this way it is distinguished from a true good-making feature such as integrity. On the theory under consideration, therefore, an intrinsically good state of affairs is likened to a property which is constitutive of the worlds which possess it.

Note that such an account of intrinsic goodness by itself yields no theory of intrinsic betterness.

But if goodness is explained in terms of good-makingness, it might be thought that betterness is to be explained in terms of better-makingness. On such a view,  $p$  would be held intrinsically better than  $q$  just in case  $p$ 's being true in a world and  $q$ 's being true in a world are paired as better-making and worse-making respectively. And by parity of reasoning with the above, one property is better-making than another for a given type of object and value when their respective possession is always at least partly constitutive of value differences which exist between superior possessors of the better-making property and inferior possessors of the worse. Thus, if intelligence is a better-making property of man than is stupidity, and if Mr Smith who is intelligent is a better man than Mr. Jones who is stupid, then some part of Smith's superiority to Jones will consist in the preferability of intelligence to stupidity.

The theory as so far developed has its attractions, especially in its account of intrinsic goodness. The treatment of intrinsic betterness is, on the other hand, considerably less appealing, especially since it seems incompatible with the supposition that one of two logically compatible states of affairs may be intrinsically better than the other. The role assigned to better-makingness is, as we have seen, to explain the extent to which things differ in value. But if properties  $P_1$  and  $P_2$  are compatible, so that two objects of unequal value may possess *both*, then the value difference between such objects can hardly be constituted—even in part—by the superiority of  $P_1$  to  $P_2$ . For example, we cannot say that Mr. Jones is a better man than Mr. Smith because Jones is moral and Smith is stupid, for it may also be the case that Jones is stupid and Smith moral. For the same reason we could not say that being a world in which Jones and Smith are both happy is better-making than being a world in which Jones is happy, and so we could not say that the former

state of affairs is intrinsically better than the latter. Yet we have a strong and natural inclination to say just this.

(4) The fourth theoretical proposal—that of Gilbert Harman—is like the second in its assumption of a numerical measure of intrinsic value, according to which these values can be subjected to ordinary mathematical operations of adding, subtracting, multiplying, averaging, etc.<sup>5</sup> Harman begins by noting that traditional explanations tend to distinguish the intrinsic from the overall value of a proposition  $p$  by holding that the former value pertains only to  $p$ 's entailments while the latter concerns every proposition whose truth  $p$  influences. This traditional way of drawing the distinction suggests the possibility of certain simple functions for measuring both types of value. Harman's first candidate for the overall value function is this:

$$(H1) \quad V(p) = \sum_q [I(q) \times PM(q/p)]$$

—that the overall value of any proposition  $p$ —i.e.,  $V(p)$ —equals the sum for every proposition  $q$ —i.e.,  $\sum_q [\dots]$ —of all the products obtained by multiplying the intrinsic value of  $q$ —i.e.,  $I(q)$ —times the degree to which  $p$  makes  $q$  more or less likely—i.e.,  $PM(q/p)$ . The analogous candidate for the intrinsic value function would presumably be a version of H1 which restricts the summing operation to propositions  $q$  which  $p$  entails—i.e.,  $\sum_{p \Rightarrow q} [\dots]$ —and eliminates the probabilistic conditioning:

$$(H2) \quad I(p) = \sum_{p \Rightarrow q} [I(q)]$$

—that the intrinsic value of any proposition  $p$  equals the sum of the intrinsic values of all the propositions which it entails.

Despite their simplicity, these two functions suffer from a serious problem, viz., they require the same values to be taken into account too many times. In illustrating this predicament for H2, let  $c$  be the intrinsically neutral state of affairs in which the cat is on the mat and  $d$  the similarly neutral state of affairs in which the dog is at the door. It seems intuitively clear that the following propositions, each of which is entailed by the final one, all have the same positive intrinsic value:  $j_1$ ,  $j_1 \& c$ ,  $j_1 \& d$ , and  $j_1 \& (c \& d)$ ; and that each gets its intrinsic value (equal to  $i$ ) because each has  $j_1$  as its entailment and because the intrinsic value of  $j_1 = i$ . But by H2 the intrinsic value of the final proposition  $j_1 \& (c \& d)$  must include the sum of all the earlier ones, so that the value of  $j_1$  would be

<sup>5</sup> Gilbert Harman, "Toward a Theory of Intrinsic Value," *The Journal of Philosophy*, vol. 64 (1967), pp. 792–804.

taken into account at least four times—instead of once—in the calculation.

Harman suggests that this difficulty might be obviated by replacing the  $I(p)$  function on the right of H1 (and H2) with a different, more specialized function which assigns what he calls a *basic intrinsic value* ( $B(p)$ ) to every proposition. A select number of propositions will receive a *positive* or *negative* basic intrinsic value while all others—including many or most intrinsically good or bad states—will not. The important requirement is that these basic values be distributed in such a way that the emended formulae:

$$(H_3) \quad V(p) = \sum_q [B(q) \times PM(q/p)] \quad \text{and} \\ (H_4) \quad I(p) = \sum_{p \Rightarrow q} [B(q)]$$

give intuitive results. Value redundancy would be eliminated, for example in the above case, by assigning  $j_1$  a positive basic value (equal to its total intuitive intrinsic value) while assigning  $j_1 \& c$ ,  $j_1 \& d$ ,  $j_1 \& (c \& d)$  basic values of zero.

This idea is ingenious, and as I shall try to show in the next section, fruitful. Yet despite these merits the theory is vulnerable to the objection that there is no plausible way to apply it so as to give intuitive results for certain propositions. Consider, for example, such a simple disjunction as  $j_1 \vee j_2$ . The states which intuitively account for its goodness (viz.,  $j_1$  and  $j_2$ ) are not among its entailments and therefore by H4 cannot enter into the calculation of its intrinsic value. Thus a basic goodness will have to be assigned either to the disjunction itself or to some proposition which it entails. But whichever proposition is assigned basic value to account for the goodness of  $j_1 \vee j_2$ , it seems that there will always be still wider disjunctions (perhaps  $(j_1 \vee j_2) \vee j_3$ ) whose goodness will need further explanation. Of course one can go on indefinitely assigning basic values to wider and wider disjunctions, but only at the price of enormous theoretical complications. This is because (a) each of the wider disjunctions (e.g.,  $(j_1 \vee j_2) \vee j_3$ ) is entailed both by various narrower ones (e.g.,  $j_1 \vee j_2$ ) and by various  $j_n$ 's (e.g.,  $j_1$ ) and (b) H4 requires that the basic value of a wider disjunction apply to the narrower ones which entail it. So in deciding to assign a basic value to explain the intuitive value of the wider propositions we shall have to reduce the basic value assigned to the narrower ones to preserve their intuitive value. And since there is no limit to the wider disjunctions which will need basic value it will be theoretically impossible to decide which basic values to assign the narrower ones.

Even apart from this difficulty it should be clear that the resulting complexity and nonintuitiveness of the system would be overwhelming.

## II

In this concluding section I shall offer a theory which incorporates a variant of Harman's function H4, but which is less problematic with respect to propositions such as  $j_1 \vee j_2$ .

Two central terms of this theory are "evaluatively prior" and "evaluatively posterior." Intuitively speaking,  $p$  is evaluatively prior to  $q$  and  $q$  is evaluatively posterior to  $p$  if knowledge of the value of  $q$  depends upon prior knowledge of the value of  $p$ . For example,  $j_1$  is evaluatively prior to both the conjunction  $j_1 \& b_1$  and to the disjunction  $j_1 \vee b_1$ , while the latter two states are evaluatively posterior to  $j_1$ . Note that in his contrast between basic and nonbasic intrinsic value, Harman seems to make tacit use of the idea that some states are evaluatively posterior to certain of their entailments but not to others. In what follows, an equally important distinction will be based on whether or not a proposition is evaluatively posterior to states *other than* its entailments.

A third primitive term of the theory is "evaluatively basic proposition" (or "basic proposition" for short). Basic propositions constitute the starting point of the system of evaluations. The intrinsic values initially assigned to them determine the value of every other proposition. Hence there is a tenuous analogy between their role in this theory and the role of atomic sentences in semantic theory. There is also, of course, a connection between this and Harman's use of "basic"—but the two uses are not the same. Here there is no such thing as "basic intrinsic value"—there are merely propositions which are first or "basic" in the assignment of intrinsic values. The value to be assigned to a basic proposition is always its full intrinsic value. The various  $j_i$ 's and  $b_i$ 's are among the most likely candidates for basic propositions, while truth functions such as  $\neg j_i$ ,  $b_i \vee j_i$ , and  $b_i \& j_i$  are almost certainly nonbasic. The precise constraints upon and purpose of the selection of basic propositions will become clearer as the theory is presented.

Six definitions derive from these and other familiar notions.

(D1)  $p$  is *evaluatively determinate* (*determinate*, for short)

iff

$p$  is not evaluatively posterior to any proposition not entailed by  $p$ .

That is, if a determinate proposition is evaluatively posterior to any proposition at all it is such only to propositions which it entails. For example, if  $j_1$  &  $b_1$  is evaluatively posterior only to  $j_1$  and to  $b_1$ , then since it is posterior only to propositions which it entails it is determinate. However, if  $j_1 \vee j_2$  is posterior to  $j_1$ , then it is posterior to a proposition which it does not entail. Hence  $j_1 \vee j_2$  would be regarded as indeterminate:

(D2)  $p$  is *evaluatively indeterminate* (indeterminate, for short)

iff

$p$  is evaluatively posterior to some proposition not entailed by  $p$ .

Note that, as in the case of  $j_1 \vee j_2$ , indeterminate propositions may be posterior to a proposition (e.g.,  $j_1$ ) which entails them; or, as in the case of  $(j_1 \& b_1) \vee j_2$ , to a proposition (e.g.,  $j_1$ ) which intersects but does not entail them. The relevance of the determinate-indeterminate distinction to earlier theories is best seen in the two extreme cases. The Rescher-type view appears to treat all propositions, other than unit sets of possible worlds, as if they were indeterminate, i.e., posterior in value only to nonentailed states. Harman, on the other hand, seems to treat every state of affairs as determinate—as if its intrinsic value were determined solely by the values of the states it entails.

The remaining definitions concern the basic propositions. These propositions, I shall claim, come in "families."

(D3)  $F$  is a *family* of basic propositions

iff

$F$  is a set of propositions all of which are basic, and the propositions in  $F$  are mutually exclusive and jointly exhaustive.

Any pair of basic propositions of the form  $p$  and  $\neg p$  would constitute the most rudimentary kind of family. (It may well be, however, that no proposition and its negation are *both* basic.) The  $j_i$ 's, when supplemented with the proposition that Jane does not exist, constitute a more complex type of basic

family. The very important theoretical role of such families is indicated in the final three definitions.

(D4)  $p$  is *indifferent* with respect to a given family  $F$  of basic propositions.

iff

$p$  intersects each member of  $F$  or  $p$  intersects no member of  $F$ .<sup>6</sup>

(D5)  $p$  is a *maximal conjunction* of the members of families  $F_1, F_2, \dots, F_n, \dots$ <sup>7</sup>

iff

$p$  is a consistent conjunction of members from  $F_1, F_2, \dots, F_n, \dots$  such that any conjunction of  $p$  with any additional member of these families is contradictory.

(D6)  $q$  is a *value constituent* of  $p$

iff

$q$  is logically compatible with  $p$ , and  $q$  is a maximal conjunction of members of those families with respect to which  $p$  is *not* indifferent.

In other words, to find the value constituents of  $p$  one must first locate the families of basic propositions with respect to which  $p$  is not indifferent, next determine the maximal conjunctions of their members, and finally eliminate those conjunctions incompatible with  $p$ . What remains are  $p$ 's value constituents—states intimately related to the value of  $p$ .

The following assumptions are those most immediately needed to support the fundamental principles of the envisaged theory. Some seem to flow directly from natural intuitions about evaluative priority and posteriority. Some, however, may be viewed as indicating those constraints (on the selection of basic propositions) which are needed to make the theory work.

The first assumption gives the logic of the relations of evaluative priority and posteriority.

(A1) Evaluative priority and evaluative posteriority are transitive, non-symmetric, and non-reflexive relations; moreover,  $p$  is evaluatively prior to  $q$  iff  $q$  is evaluatively posterior to  $p$ .

<sup>6</sup> It will be convenient later to amend this definition slightly. See fn. 9.

<sup>7</sup> In this use of "conjunction" any number of propositions may be conjoined—the conjunction of a single proposition  $p$  is simply to be identified with  $p$  itself. The notation " $F_1, F_2, \dots, F_n, \dots$ " is meant to indicate that any number of families may be involved.

The second assumption describes the structure which is required of the array of basic propositions.

- (A2) Each basic proposition is a member of a family of basic propositions, and the families of basic propositions are such that no basic proposition entails any other.

Notice that the logical nature of a family entails that the maximal conjunctions of the members of one or more given families (i.e., the value constituents of states relevantly related to these families) are mutually exclusive and jointly exhaustive, a feature important to the definition of value in indeterminate propositions.

The third assumption indicates where the relations of evaluative posteriority and priority occur.

- (A3)  $p$  is evaluatively posterior to  $q$  iff (a)  $p$  is non-equivalent to  $q$ , and  $q$  is a value constituent of  $p$ , or (b)  $p$  is non-equivalent to  $q$ , and  $q$  is a basic conjunct of one of the value constituents of  $p$ .

These definitions and assumptions yield the following two theorems and so provide a simple way of distinguishing determinate from indeterminate states.<sup>8</sup>

- (T1) A proposition is evaluatively determinate iff it has one and only one value constituent or it has no value constituent.

That is, determinate propositions either have no constituents at all because they are indifferent to every family of basic propositions (e.g., the hat is on the mat or  $2+2=4$ ) or they are completely contained within a single constituent (e.g.,  $j_1$  &  $b_1$ ).

- (T2) A proposition is evaluatively indeterminate iff it has two or more value constituents.

A proposition is indeterminate simply because it overlaps any one of its value constituents. For example,  $(j_1 \vee j_2)$  has both  $j_1$  and  $j_2$  as its constituents, but entails neither.

The value topography generated in this system suggests the following principles for determining value.

- (P1) The intrinsic value of any consistent conjunction composed exclusively of basic propositions as conjuncts  
=the sum of the intrinsic values of the conjuncts.

Since all value constituents are conjunctions of this sort, P1 gives a simple method for evaluating each value constituent of any proposition. Given this, determinate propositions may be treated as follows:

- (P2) The intrinsic value of an evaluatively determinate proposition  $p$  having no value constituents  
=0.  
(P3) The intrinsic value of an evaluatively determinate proposition  $p$  with one value constituent  
=the intrinsic value of that value constituent.

The intrinsic value of any determinate proposition (including the basic ones) thus turns out to be the sum of the values of the basic propositions it entails.

Indeterminate propositions present, as has been apparent all along, a special problem. The solution which I favor and shall presently defend is to avoid the task of determining a unique numerical value for each indeterminate proposition as such, but instead to relativize the value of such propositions to their member worlds. Thus, I shall not speak of the value of an indeterminate proposition in itself, but rather of its value *in a given possible world  $w$*  in which it is true.

- (P4) The intrinsic value of an evaluatively indeterminate proposition  $p$  in a world  $w$  in which  $p$  is true  
=the intrinsic value of the value constituent of  $p$  true in  $w$ .

Since one and only one constituent of  $p$  will be true in a given world, P4 yields a unique  $p$ -value for each world in  $p$ . P4 also gives rise to a clear and intuitive test of the intrinsic goodness, badness, or neutrality of indeterminate propositions.

- (P5) An evaluatively indeterminate proposition is good iff it is good in each of its possible worlds; bad iff it is bad in each of its pos-

<sup>8</sup> The proof of T1 and T2 is sketched as follows: From A2 and D3-D6, it follows that a proposition with one or fewer value constituents entails all of its constituents and all the basic conjuncts which make them up. Hence, by A3, such a proposition is posterior only to states which it entails i.e., is determinate. On the other hand, if  $p$  has two or more value constituents it must, by D6, be logically compatible with each of them; and so, since  $p$ 's value constituents must be mutually exclusive,  $p$  cannot entail any of them. Thus, by A3,  $p$  is evaluatively posterior to states which it does not entail, i.e., is indeterminate. On assumption A1, T1, and T2 follow directly.

sible worlds; and neutral iff either it is neutral in each of its possible worlds, or not good in each and not bad in each.

Critical examination of the theory resolves into two specific questions: First, how plausible are the constraints which the theory places on the selection of basic propositions? That is, do the propositions we would intuitively think of as basic fit the theory? And second, how plausible are the results obtained for nonbasic propositions? With regard to the first question, it appears that—given one minor accommodation—the constraints on the basic propositions are quite acceptable. It is natural to suppose that the most evaluatively prior of all states of affairs are those which locate a specific sentient individual at a specific point along an evaluatively relevant dimension such as happiness, virtue, wisdom, etc. Thus for each pair consisting of an individual and a dimension there will be a distinct basic proposition for each point on that dimension which that individual may occupy. All such propositions (for a given individual and dimension) will be disjoint and together they will exhaust all possibilities other than the individual's non-existence. Thus, if the appropriate negative existential is added to each such set, the result will be a family.<sup>9</sup> (The "minor accommodation" mentioned above consists precisely in this addition.) The supposition that no basic proposition entails any other certainly holds for all members of the same family. That it holds between members of different plausible families (e.g., between  $j_1$  and  $b_1$ ) is also defensible. While there may be all sorts of contingent, perhaps even physically necessary, connections between Jane and Bill's happiness, or Jane's virtue and Jane's happiness, etc., it seems sensible to suppose that none of these connections is strictly necessary.

In addition, an advantage of the constraints of the present theory is that they permit the application of some of the earlier semantic theories to the basic propositions. For example, since the members of any pair of basic propositions are either logically disjoint or logically independent, the "maximal similarity" version of Theory 1 might be used to account for comparisons of value among them.

Moreover, the assignment of basic propositions to the absolute categories can be left to the "good-making" account of Theory 3. It would seem that a theory which traces all value back to a limited set of basic propositions may leave the semanticist with a more manageable task.

Assuming that the constraints on the selection of basic propositions are more or less plausible, can the same be said for the results projected by Pr-P5 onto the nonbasic propositions? We may illustrate these results for the simplified case in which Jane and Bill are the only individuals and happiness the only relevant dimension. Hence, the basic propositions are

...  $j_{-n}, \dots, j_{-1}, j_0$ , Jane does not exist,  
 $j_1, \dots, j_n, \dots$   
 ...  $b_{-n}, \dots, b_{-1}, b_0$ , Bill does not exist,  
 $b_1, \dots, b_n, \dots$

Let us further assume, for convenience, that the value of each basic proposition corresponds to the degree of happiness involved, and that in this world Jane is happy<sub>2</sub> and Bill is happy<sub>-1</sub>. Consider the results for the following sample propositions:

- (1)  $j_1$  &  $b_1$ : Determinate with constituent  $j_1$  &  $b_1$ ; Intrinsic value = 2.
- (2)  $j_1$  & grass is green: Determinate with constituent  $j_1$ ; I.V. = 1.
- (3) A proposition containing a single world in  $j_1$  &  $b_1$ ; Determinate with constituent  $j_1$  &  $b_1$ ; I.V. = 2.
- (4) Grass is green: Determinate with no constituent; I.V. = 0.
- (5)  $j_1 \vee j_2$ : Indeterminate with constituents  $j_1$  and  $j_2$ ; I.V. in this world = 2; Intrinsic Value Status: Good.
- (6)  $\neg j_1$ : Indeterminate with constituents consisting of all the degrees of Jane's happiness other than  $j_1$ ; I.V. in this world = 2; I.V.S.: Neutral.
- (7)  $j_1 \vee b_2$ : Indeterminate with constituents consisting of all the maximal conjunctions of  $b_i$ 's and  $j_i$ 's which include either  $j_1$  or  $b_2$ ; Not true in this world; I.V.S.: Neutral.
- (8)  $\neg j_1 \vee b_2$ : Indeterminate with constituents

<sup>9</sup> Notice that if there is more than one family of basic propositions for Jane (e.g., a "happiness" family and a "virtue" family), then these families will share the basic proposition which denies Jane's existence. Thus, a proposition such as  $j_1$  will not be indifferent to the "virtue" family for Jane since there is one member of that family (Jane's nonexistence) which it does not intersect. Note also that such neutral states as that Jane is a girl are not indifferent to any of these families—again because of their nonintersection with the negative existential. Both these results are awkward. The awkwardness is eliminated, however, if the definition of "indifference" is revised such that a proposition is indifferent to a family iff it either intersects all the members of the family, intersects all members other than the negative existential, or intersects no members of the family.

consisting of all the maximal conjunctions of  $b_i$ 's and  $j_i$ 's which include either  $b_2$  or any  $j_i$  such that  $i \neq 1$ ; I.V. in this world = 1; I.V.S.: Neutral.

- (9) That Jane is positively happy to some degree or other; Indeterminate with constituents consisting of all the positive  $j_i$ 's; I.V. in this world = 2; I.V.S.: Good.

So far as my intuitions go, these results are plausible. (7), perhaps, is a slight exception. But even that result should, I think, be accepted. For  $j_1 \vee b_2$ , unlike  $j_1 \vee j_2$ , allows for the possibility that Jane is unhappy and is compatible with such bad propositions as  $j_1$  &  $b_{-10}$  and  $j_{-10}$  &  $b_2$ . Thus, even though its disjuncts are both good, the fact that they are not disjoint from one another permits relevant bad as well as good realizations of their disjunction. Hence its neutrality.

Various schemes are possible for assigning precise numerical values to the indeterminate propositions. For example, those having a finite number of constituents may be assigned the average value of their constituents. Thus, the value of  $j_1 \vee j_2$  would be midway between that of  $j_1$  and  $j_2$ . Schemes are even possible for processing indeterminate propositions having an infinite number of value constituents—although I know of no plausible scheme which handles all such propositions. These procedures, however, involve one in what may be needless complication. For the point of a theory of intrinsic value is first to secure strongly felt intuitions about the intrinsic values of various states of affairs (e.g. the intuition that  $j_1 \vee j_2$  is intrinsically good) and second to service the theory of overall value.

That we have no strong intuitions about the precise values of most indeterminate propositions seems clear enough. If we have any intuitions at all,

they are restricted to indeterminate propositions having a finite number of constituents, and thus may be accommodated by the simple averaging technique indicated above. More important, calculations of *overall* value may safely ignore the precise intrinsic value of indeterminate propositions. The overall value of a proposition  $p$  is best regarded as the sum of the intrinsic values of all the basic propositions it entails plus its extrinsic value, where its extrinsic value is the sum of the products of the intrinsic value of each *basic* proposition  $b$  (which  $p$  does not entail but which is logically compatible with  $p$ ) times the amount that  $p$  makes  $b$  more or less likely.<sup>10</sup> Thus, the only intrinsic values needed to solve the problem of overall value are those of the basic propositions.

In summary, the distinctive features of the theory just presented are (1) the selection of a relatively limited subset of propositions as basic, (2) the division of propositions into the categories of determinate and indeterminate, and (3) a method for determining the value constituents of each proposition. The value status of a constituent is held to equal the sum of the values of the basic propositions it entails; and the value status of each proposition is determined by the values of its constituents. I have claimed that the results of such a system are plausible, and that the constraints placed on the selection of basic propositions appear to make them conformable to earlier theories of value. Of course, the matters involved are very complex and several important problems have not been discussed—including whether the theory can explain the phenomenon often described as the "defeat" of intrinsic goodness or badness. While I am confident that it can meet both this and other challenges, these are matters for a further discussion.<sup>11</sup>

*University of California, Los Angeles*

*Received March 15, 1973*

<sup>10</sup> That is, more or less likely than it would be on the disjunction  $p \vee \neg p$ . This differs from Harman's formula in two major respects. First, the reference to basic value is replaced by a reference to basic propositions and second, the *total* intrinsic value of determinate propositions is included as an addend in calculating their overall value. A curious result of Harman's formula H3 (avoided by my alternative) is that a determinate proposition having no influence on propositions contingently related to itself will nevertheless have an overall value *lower* than its intrinsic value (see Harman, *op. cit.*, p. 799).

<sup>11</sup> I wish to thank Tony Anderson, John Bennett, Tyler Burge, David Lewis, and David Kaplan for extremely helpful comments and suggestions about earlier drafts of this paper.

## VI. PARADOXICAL STATEMENTS

GARETH MATTHEWS

The disappearance [during the printers' strike] of the *Times* Sunday book section only made us aware that it had never existed.

—EDMUND WILSON

But these cares and anxieties were the only kind of happiness possible for Dolly.

—LEO TOLSTOY

IF we take "formal contradiction" to mean *something that has the explicit form of a contradiction* (such as "It's raining and it isn't raining"), we can safely say that not many paradoxes are stated as formal contradictions. It may be initially tempting to divide the remaining paradoxical statements into

- (a) those that, although they are not themselves formally contradictory, entail statements that are

and

- (b) those that, although they are not themselves formally contradictory and do not even entail other statements that are, nevertheless contain terms that normally function as either contraries or contradictories.

An example of the first sort would be this opening line from Dickens' *A Tale of Two Cities*:

- (1) It was the best of times, it was the worst of times.

Although (1) is not a formal contradiction it entails

- (2) It was the best of times and it was not the best of times

which is.

An example of the second sort would be the aphorism,

- (3) In winning the battle he lost the war.

(3) is not formally contradictory; neither does it entail anything formally contradictory. It is paradoxical simply because (so the story would go) it couples two terms (viz., "win" and "lose") which are normally contraries.

But this story will not do. It leads us to expect that

(4) Sometimes you win, sometimes you lose would be a paradox; for it, too, couples the normally contrary terms, "win" and "lose." But clearly (4) is not a paradox.

I am inclined to think we won't understand why (3) is a paradox and (4) is not one unless we can offer an analysis which brings (3) more into line with (1). I shall now offer such analysis.

But first I must introduce the notion of formal contrariety, to supplement the notion of formal contradiction. The sentence, "All Bavarians are boisterous," is a contrary of the sentence, "Some Bavarians are quiet and querulous." For, although statements made with these sentences may both be false, they can't both be true. (They will both be false if indeed some Bavarians are quiet, but those that are quiet are not also querulous.) The contrariness of "All Bavarians are boisterous" and "Some Bavarians are quiet and querulous" is not, however, a purely formal matter; it rests on the meanings assigned to the categorematic terms, "boisterous," "quiet," and "querulous." These sentences are, as we might say, material contraries.

By contrast, the sentences, "All Bavarians are boisterous" and "Some Bavarians are neither boisterous nor brave," are formal contraries. Again, the statements made with these sentences may both be false, but can't both be true. But their contrariness is a purely formal affair. It is independent of the meanings assigned to the respective categorematic terms, so long as the same meaning is assigned to each occurrence of a single term. For the sake of brevity I shall use "is formally incompatible with" to mean "is either the formal contradictory or a formal contrary of."

Now a paradox (I want to say) is

- (D) a statement made with a sentence which is formally incompatible with some sentence that can be taken, on a standard reading of its terms, to express a conceptual truth.

It might seem that, when we consider sentences that can be used to express conceptual truths, the



notion of contrariety becomes inapplicable. After all, what marks off contraries from contradictories is the fact that a pair of contraries (but not a pair of contradictories) can both be false. But surely no conceptual truth can be false.

This objection rests on a confusion. It is sentences I am calling formal contraries. On one assignment of meaning to its categorematic terms, "All chairmen are male" will express a conceptual truth; on that assignment a statement made with it will be such that it can't be false. But other assignments are possible. On some of these the resulting statement will be false; therefore "All chairmen are male" is merely the contrary of the sentence, "Some chairmen are neither male nor maudlin."

So, according to (D), a paradox is a statement made with a sentence formally incompatible with some sentence that can be used to express a conceptual truth.

Let's call a sentence that expresses a relevant conceptual truth "a background sentence" or "the background sentence" and the paradoxical sentence itself "the foreground sentence." The background sentence and the foreground sentence stand in a relation of either formal contradiction or formal contrariety. But in accordance with a rule of interpretation which I shall try to ferret out in a moment, we know to understand some expression in the foreground sentence in a sense, or with a qualification, different from the sense, or qualification, (if any), we assign to that same expression in the background sentence.

Perhaps examples might help make the analysis clearer. Consider the paradox,

- (5) Some inexpensive cars are not inexpensive.

One might take this as the background sentence:

- (6) All inexpensive cars are inexpensive.

Or this:

- (7) Everything that is inexpensive is inexpensive.

(7) is a formal contrary of (5); (6) is the formal contradictory of (5). Let's use (6). By a rule of interpretation yet to be uncovered, we know to understand (5) in such a way that the statement made with (5) is not really incompatible with the conceptual truth expressed by (6). Most plausibly (5) is a way of saying that some cars that are inexpensive to buy are not inexpensive to buy *and* maintain, or perhaps that some cars that are inexpensive relative to other cars are nevertheless not inexpensive *tout court*.

Let's return again to the Dickensian line,

- (1) It was the best of times, it was the worst of times.

The background sentence that goes with it is either its contradictory,

- (8) Either it isn't the best of times or it isn't the worst of times

or, perhaps better, this contrary:

- (9) Nothing is both the best of times and the worst of times.

Certainly (9) can be used to express a conceptual truth; and it is a formal contrary of (1). But we know to understand (1) in some such way as this:

- (10) In some respects it was the best of times and in other respects it was the worst of times.

(9) and (10) are not, of course, incompatible—formally or informally.

The conceptual truth suggested by (3) is perhaps the most interesting of the lot; it is something that might be expressed by this background sentence:

- (11) It isn't in winning something that one loses something.

(11) is ambiguous. To make it express a conceptual truth one reads it to mean something like this:

- (12) Winning something is never in itself losing something.

The other reading of (11) is this:

- (13) Winning something never has the result that one loses something.

(13), far from expressing a conceptual truth, expresses a falsehood.

I have used the notion of a conceptual truth in a rather offhand way. My offhandedness has been deliberate. The notion of a paradox is a rather commodious one. My account of paradox will not have much chance of success unless I can make use of a commensurately commodious notion of conceptual truth.

Whether our notion of conceptual truth is, or should be, commodious enough to accredit conceptual truths to go with pragmatic paradoxes like "I don't exist," "I'm not here," and "I don't speak English" is an interesting question. No doubt some philosophers would be happy to say that "I exist" can be used to express a conceptual truth. Fewer would want to say the same thing about

"I'm here." Perhaps nobody thinks that "I speak English" can be used to express a conceptual truth.

I suspect it would be better to say that pragmatic paradox is not really a variety of paradox; rather, it is a variety of seeming paradox, or near-paradox. One might go on to explain that a pragmatic paradox is a statement made with a sentence that is formally incompatible with some sentence, *s*, such that, if *s* were used to make an assertion, the mere fact the assertion had been made would itself afford a good reason for thinking that what was asserted was true. To develop and clarify this suggestion I should need to deal with the family of notions that have grown up in recent literature on the so-called performative aspect of Descartes' *cogito*. Instead of undertaking that project I shall turn to another class of near-paradoxes.

My account of paradox is certainly not guaranteed to fit what is sometimes called "condensed paradox" (cf. Richard D. Lanham, *A Handlist of Rhetorical Terms* [Berkeley, 1969], p. 70), namely, oxymoron. But that is all to the good. For an oxymoron is not, as such, a paradox. Nor is the statement one makes with an oxymoron, as such, a paradox. An oxymoron is a phrase that places opposing terms in close syntactic proximity ("gerafftente syntaktische Verbindung," Heinrich Lausberg calls it in his *Handbuch der Literarischen Rhetorik* [Munich, 1960], vol. 1, p. 398).

Typically an oxymoron is made up of a modifier and a term modified such that the two terms are either contraries or contradictories of each other,

or, if the modifier is an adverb, its adjective form is a contrary or contradictory of the adjective or verb it modifies. Thus "cruelly kind," "beautifully hideous," "makes haste slowly," "a true lie," and "a hardtop convertible" are examples of oxymora.

When, as is typical with oxymoron, one of the opposing terms modifies the other, this fact in itself means that there is no real contradiction of a conceptual truth, only the semblance of such contradiction. Thus it might seem that

(14) Feruccio's dress is fastidiously casual

is a genuine paradox, on the grounds that this contrary of (14),

(15) Nothing is fastidiously casual

expresses a conceptual truth. But (15) doesn't express a conceptual truth, though there is a sentence easily confused with (15) which does, namely, this:

(16) Nothing is both fastidious and casual.

So (14) contradicts (15) all right, but it doesn't contradict (16); and it is (16), not (15), that can be used to express a genuine conceptual truth.

The somewhat more general point is that, although some oxymora of the syntactical form "*x* is *F*-ly *G*" entail the corresponding "*x* is *G*"<sup>1</sup> ("*x* is beautifully hideous" entails "*x* is hideous") and some entail the corresponding "*x* is *F*" (on perhaps its most likely reading, "*x* is cruelly kind" entails

<sup>1</sup> Terence Parsons in his "Some Problems Concerning the Logic of Grammatical Modifiers" (*Synthese*, vol. 21 [1970], p. 323), suggests dividing modifiers into those that are standard and those that are non-standard, where to be standard is, roughly, to be eliminable *salva veritate*. Thus "slowly" is a standard modifier, since if "Dan drives slowly" is true, then so is "Dan drives." But "reputedly" is non-standard, since "Smith reputedly smokes" can be true though "Smith smokes" is false.

The way Parsons himself draws this distinction one should be able to say, as a perfectly general thing, whether a given modifier is standard or non-standard. But of course the best behaved of standard modifiers might turn up in an oxymoron and behave, in that context, non-standardly. Thus "slowly," a modifier Parsons counts as standard, turns up non-standardly, in this exchange:

Impatient nephew: "Where in the world is Uncle Milton? I told him to make haste."

Uncle Milton's longsuffering wife: "I'm afraid your Uncle Milton makes haste slowly."

One might, of course, hold onto Parson's way of drawing the standard/non-standard distinction by making clear that it applies only to literal uses of language; oxymoron, being a figure of speech, would then be ruled irrelevant. Alternatively one might give up the Parsons approach and require that the question "Standard or non-standard?" be relativized to a particular use of the expression in question. Then one could say that, although "slowly" (for example) is usually a standard modifier, it may be non-standard in the context of an oxymoron.

No doubt one thing that recommends the once-and-for-all approach is the hope it offers of enabling us to tell from an appropriate dictionary entry for (say) "slowly" that the inference from "*x* *G*'s slowly" to "*x* *G*'s" is a good inference. But presumably the relevant part of the dictionary entry would have to be based on literal uses of "slowly." To know whether the inference in a case before us is good we should need to know whether the case before us involves a literal use of language. And one might be sceptical about whether it is reasonable to think a dictionary could be constructed that would enable us to decide in an automatic fashion whether a given expression is an oxymoron, let alone to decide in an automatic fashion whether a given expression is being used figuratively in some way or other, or anyhow being used figuratively in some way that might affect the inference from "*x* *G*'s slowly" to "*x* *G*'s."

" $x$  is cruel") none entails " $x$  is both  $F$  and  $G$ ." Yet the relevant conceptual truth would have to be "Nothing is both  $F$  and  $G$ ." And that is not contradicted by " $x$  is  $F$ -ly  $G$ ."

An even more general point could be constructed along similar lines to show that, typically anyway, oxymora of other syntactic forms do not yield genuine paradoxes either.

I say my account of paradox is meant to be perfectly general. To be more careful I should say that it is meant to be a general account of paradox *as figure of speech*. What in philosophy is perhaps most often referred to as a paradox is not a figurative statement at all; rather it is *the conclusion that  $p$* , where a suitable replacement for ' $p$ ' expresses an apparently absurd or contradictory consequence of some otherwise seemingly acceptable notion or theory or statement.<sup>2</sup> Zeno's Paradoxes, the Paradoxes of Confirmation, the Paradox of Sovereignty, the Paradox of Analysis, the Paradox of Omnipotence, the Liar Paradox, and Russell's Paradox are all examples of paradox in that sense of the word. My account of paradox as figure of speech is not directly relevant to those paradoxes, though it may be indirectly relevant. To explain how it may be indirectly relevant I shall say a word about Zeno.

Presumably Zeno put forward his paradoxes to show the untenability of the notion of plurality, and perhaps also the notion of movement.<sup>3</sup> We have no reason to think that Zeno meant his conclusions to be taken figuratively. But one could, of course, use a Zenonian sentence, e.g., this one,

(17) The arrow is at rest throughout its flight

to make a figurative statement. Its background sentence would be, perhaps, this truism:

(18) Nothing is at rest throughout its flight.

(18) can be used to express a conceptual truth. But we should know to find an equivocation between the foreground and background sentences (so interpreted). "Throughout its flight" in (17) might be taken to mean "at every moment during its flight" and in (18) to mean "for every period during its flight."

I repeat, there is no reason to think that Zeno himself meant to be making figurative statements. On the contrary, there is every reason to think otherwise. But if one were to use one of his conclu-

sions to make a figurative statement, the analysis I have offered would seem to fit the result.

\*            \*            \*

If my account of paradox as figurative statement is more or less correct, we know to take the foreground sentence of a paradoxical statement in two different ways—first in the way that suggests an appropriate background sentence (this reading is rather straightforward and unqualified), and then in another way (perhaps with a special sense or qualification assigned to some key expression) to keep what is expressed by the foreground sentence from colliding with the conceptual truth expressed by its background sentence. Thus we take "He was wise and he was foolish" in a straightforward way that would enable us to derive the background sentence, "Nobody is both wise and foolish," and then we take it less straightforwardly for what we suppose the speaker to have meant by it. Let's call the practice of taking the foreground sentence in the second way "reconstructing the foreground sentence."

I promised to say something about the rule of interpretation according to which we reconstrue the foreground sentence. I shall present two candidates for such a rule, the more radical one first.

The more radical candidate is made up of two parts, viz., a Nonsense Principle and a No-Nonsense Principle.

#### *Nonsense Principle:*

A sentence formally incompatible with a sentence that expresses a conceptual truth doesn't make any sense unless it is reconstrued to avoid the incompatibility.

#### *No-Nonsense Principle:*

Other things being equal, the declarative sentences of a speaker or writer should be interpreted so that they make sense.

The force of "other things being equal" in the No-Nonsense Principle is to demand a justification for considering declarative sentences nonsense. One suitable justification might be that the speaker is insane. Another might be that the speaker is composing nonsense verse or giving an example of nonsense.

Consider again the opening line from *A Tale of Two Cities*,

<sup>2</sup> Compare W. V. Quine's rough characterization of paradox—"... a paradox is just any conclusion that at first sounds absurd but that has an argument to sustain it," *The Ways of Paradox* (New York, 1966), p. 3.

<sup>3</sup> But see G. E. L. Owen, "Zeno and the Mathematicians," *Proceedings of the Aristotelian Society*, vol. 58 (1957-58), pp. 199-222.

- (1) It was the best of times, it was the worst of times.

According to the No-Nonsense Principle, (1) should be interpreted (other things being equal) so that it makes sense. But according to the Nonsense Principle it won't make sense unless it is reconstructed. So we reconstrue it to mean,

- (10) In some respects it was the best of times, in other respects it was the worst of times.

In their paper, "In Defense of a Dogma" (*Philosophical Review*, vol. 65 (1956), pp. 141-158), P. F. Strawson and H. P. Grice appeal to something like the Nonsense Principle to distinguish between "My neighbor's three-year-old child understands Russell's Theory of Types" and "My neighbor's three-year-old child is an adult." In each case we try to construe the utterance non-literally. (The first utterance looks like hyperbole, the second like paradox.)

In the first case Grice and Strawson suggest we might respond to the speaker by saying, "You mean your child is a particularly bright lad." If the speaker rebuffs this and similar efforts to construe his words nonliterally, we simply disbelieve him. "But," say Grice and Strawson, "if the child were then produced, and did (as one knows he would not) expound the theory correctly, answer questions on it, criticize it, and so on, one would in the end be forced to acknowledge that the claim was literally true . . ." (p. 150).

The second statement ("My neighbor's three-year-old child is an adult") sounds like paradox. Grice and Strawson suggest we respond to it with "You mean he's uncommonly sensible or very advanced for his age"; or with "Perhaps you mean that he won't grow any more, or that he's a sort of freak, that he's already fully developed." (These recommended responses are clearly efforts at what I am calling "reconstruing the foreground sentence.") If these and similar attempts at reconstruction fail, Grice and Strawson are ready to conclude that the speaker's words are simply senseless. "For unless he is prepared to admit that he is using words in a figurative or unusual sense," they insist, "we shall say, not that we don't believe him [as was appropriate in the first case], but that his words have no sense" (p. 151).

The reluctant conclusion Grice and Strawson come to in this recalcitrant case makes clear they think that it is the threat of nonsense that moves us to try to reconstrue a foreground sentence.

There are, however, good reasons to be worried about the Nonsense Principle. Typically the contradictory of an unreconstructed foreground sentence will express a necessary truth. Even if one is comfortable about rejecting the Law of Excluded Middle, there remains something queer about the idea that, whereas ' $p$ ' is nonsense, ' $\text{not } p$ ' is not only true, but necessarily so.

Then there is a worry about *reductio ad absurdum*. Consider the following sketch of an argument:

- |            |  |
|------------|--|
| To prove:  | (i) There is no greatest cardinal number.  |
| Suppose:   | (ii) There is a greatest cardinal number, namely, $n$ .  |
|            | (iii) Being a cardinal number, $n$ has a successor.  |
|            | (iv) Being a greatest number, $n$ has no successor.  |
| So:        | (v) $n$ has a successor and $n$ has no successor.  |
| Thus:      | (vi) If something, $n$ , is a greatest cardinal number, then it has a successor and it has no successor. |
| Therefore: | (vii) There is no greatest cardinal number.  |

Most people would consider the above the sketch of a valid argument. But the Nonsense Principle seems to threaten its validity. Either we can reconstrue (v) or we can't. If we can, then (v) might well come out true and the argument be invalid. If we can't, then, according to the Nonsense Principle, (v) is nonsense. That result, too, seems to wreck the validity of the argument. Surely (v) can't be conditionalized, as in (vi), unless (v) makes sense.

One thing that gives the Nonsense Principle its plausibility is the seeming ridiculousness of somebody's saying (e.g.) " $2 + 2 \neq 4$ " and then resisting our efforts to reconstrue his foreground sentence. But two comments are in order here. First, a very complicated sentence that turns out to be self-contradictory, say " $(89 + 36 - 21 + 2 - 102) \neq 4$ ," does not invite the same degree of puzzlement, perhaps does not invite puzzlement at all. And, second, a special circumstance, such as the fact that we are in the midst of a *reductio ad absurdum* argument, might encourage us—not just to allow that perhaps " $2 + 2 \neq 4$ " makes sense—but to insist that it does; what it states in a perfectly meaningful way is a necessary falsehood.

These last two comments can serve to introduce the second candidate for a rule of interpretation I

want to discuss, viz., the Detrivialization Principle. The Detrivialization Principle is an alternative to the combined Nonsense and No-Nonsense Principles.

*Detrivialization Principle:*

Other things being equal, a speaker or writer is not to be taken as trying to say or write something either trivially true or trivially false.

The "other-things-being-equal" clause with which the Detrivialization Principle begins allows for contexts in which one might want to make non-paradoxical use of a trivially false or trivially true statement, as, for example, in a formal proof.

One could look upon the Detrivialization Principle as a broader version of something W. V. Quine seems to be getting at in this passage from *Word And Object* (Cambridge, 1960):

... when to our querying of an English sentence an English speaker answers 'Yes and no', we assume that the queried sentence is meant differently in the affirmation and the negation; this rather than that he would be so silly as to affirm and deny the same thing. (P. 59.)

One attractive feature of the Detrivialization Principle is that it applies to the figure of speech, tautology (not to be confused with the logical form, tautology), as well as to paradox. Tautologies ("Business is business," "Boys will be boys," "What's done is done") have established, non-trivial, uses in our language.

Of course the examples of tautologies I have given so far are clichés; one could have learned their meanings as set phrases. But we are free to make up new tautologies with some reasonable hope that they will be taken in the right way. If I said, for example, "I want to write a book because, well ... if you've written a book you've written a book," my hearers would know to take my tautology non-trivially. They would make tacit appeal to the Detrivialization Principle.<sup>4</sup>

I should point out that a sentence which is sufficiently complicated will not fit the terms of the Detrivialization Principle. Such a sentence may be false, even necessarily false; but, being complicated, it will not be trivially false. This point suggests (wisely, I think) that we ought to amend our

definition of "paradox" so as to restrict its application to sentences that are without much logical complexity. A paradox (we shall now say) is

(D\*) a statement made with a sentence of no great logical complexity that is formally incompatible with some sentence that can be taken, on a standard reading of its terms, to express a conceptual truth.

Clearly a speaker can reasonably count on our reconstruing his sentence only when he can reasonably count on our realizing that otherwise the foreground sentence will contradict a sentence that expresses a conceptual truth *and* can reasonably count on our being confident he can be confident of our confidence. Without such iterated confidence paradoxical statements could not have the function they have. Thus a statement made with the relatively simple sentence, "He was wise and he was foolish," is a paradox. But this hardly counts as paradox:

(19) He was wise and, if either wise or foolish, then if not foolish, not wise, and if wise, not foolish.

(What (19) counts as is "exercise in logic.")

\* \* \*

My account of paradox may seem to require that we count as paradoxical many statements it is not natural to regard in this way. Suppose, for example, a father, in explaining to his son how one tells from a clock what time it is, first says something about the hour hand, then says something about the minute hand, then adds this:

(20) The third hand is the second hand.

We might plausibly credit the father with making a pun (whether intentional or not); perhaps it would be less natural to credit him with having stated a paradox.

The identification, *pun*, tends to pre-empt the identification, *paradoxical statement*, because one of the terms of (20) is clearly homonymous in such a way that (20) can be reconstrued to avoid clash with the conceptual truth expressed by

(21) No third hand is a second hand

<sup>4</sup> An interesting appeal to something like the Detrivialization Principle to interpret a tautology is to be found in the whimsical exegesis of a poem by Simonides in Plato's *Protagoras*. Concerning the tautology, "All is fair that is unmixed with foul," Socrates says this:

"He does not say this in the sense in which he might say, 'all is white that is unmixed with black'—that would be ridiculous on many counts—but meaning that for his part he accepts without censure the middle state." (346d—Guthrie, trans.)

simply by assigning to "second" another of its perfectly standard meanings.

The point is a general one; where reconstruction requires only that a key expression be given another of its perfectly standard meanings, it will be more natural to say that the speaker has made a pun than that he has stated a paradox.

Should we then tighten up (D\*) in such a way that a statement like (20) would not count as a paradox at all? We could certainly do that. But perhaps it is simpler and more satisfactory to say that paradoxes in which reconstruction is simple and straightforward in this way are *degenerate* paradoxes. They are degenerate because they rest on puns.

\* \* \*

Just as some paradoxes are such that their foreground sentences are terribly easy to reconstrue, so others are such that their foreground sentences are terribly difficult to reconstrue. In this latter group the ones of most interest to philosophers will be those the reconstruction of whose foreground sentences is, not just difficult, but philosophically problematic.

Here are three examples:

- (22) What is not, exists, as well as what is.
- (23) Those and only those are free who know that they are not free.
- (24) If a proposition tells us what the weather logically must be, then it doesn't tell us what the weather is.

The classical atomists used something like (22) to try to say that vacuum, or the void, exists, as well as atoms. In talking this way they accepted Parmenides' notion that whatever we can talk or think about exists. But, at the same time, they tried to reject Parmenides' notion that we can't talk or think about what is not. Whether the atomists can be said to have got away with their selective respect for Parmenides' notions is a question of continuing philosophical interest. In more sophisticated forms it is still with us today.

Thesis (23) seems to offer a way of saying something that the Stoics wanted to say. How difficult

(23) is to reconstrue depends on what content we are allowed to assign the final "they are not free." If we are allowed to make a contrast between (say) outward circumstance and inward attitude and conclude that it is in outward circumstance that no one is free, then reconstruction comes rather easily. But if all or part of the backing for "they are not free" is, say, a perfectly general argument for determinism, or fatalism, then one may well wonder whether any satisfactory reconstruction of (23) is possible.

I have lifted (24) from an unpublished paper by Morris Lazerowitz. It appears in a context in which Professor Lazerowitz is making the claim that the proposition, "Either it is raining or it is not raining," is not about the weather. The problem of reconstruction for (24) is intimately connected with the very difficult problem of saying what it is that propositions (or statements) "tell us."

\* \* \*

Provided we allow for there being degenerate paradoxes on one extreme, as well as profound and even philosophical ones on the other, we can say, I think, that (D\*) stands up quite well as an account of what "paradox" means in its application to paradoxical statements. We can add, I think, that the Detrivialization Principle fits the way we recognize, use, and understand paradoxical statements. In a plausible fashion the Detrivialization Principle connects the figure of speech, paradox, with the figure of speech, tautology. Unlike the Nonsense Principle, the Detrivialization Principle does not threaten the Law of Excluded Middle or the validity of *reductio ad absurdum*. Neither the Detrivialization Principle nor (D\*) requires us (wrongly) to count as paradoxes statements of great logical complexity. I conclude that (D\*) gives the sense of "paradox" relevant to our discussion and that the Detrivialization Principle codifies our use and interpretation of those figurative statements we recognize as paradoxes.<sup>5</sup>

*University of Massachusetts/Amherst*

*Received February 12, 1973*

<sup>5</sup> In writing this paper I have profited from the criticisms of Bruce Aune, Robert Coburn, Fred Feldman, Terence Parsons, and Malcolm Smith.

## VII. A THEORY OF VOLITION

CHARLES RIPLEY

### I. BASIC ACTIONS AND SKILLS

CERTAIN philosophers, most notably Arthur Danto and Roderick Chisholm, have argued that the study of human action is furthered by defining a class of basic actions, "those things," as Chisholm puts it, "which, as one might say, we do without having to do other things to get them done."<sup>1</sup> Moving one's arm, for instance, is a basic action, whereas moving a book is not. At pain of infinite regress, every action must have at least one basic action as a component. Basic action theorists take this concept to be primitive and unanalyzable and assume that all basic actions are movements of the agent's body. I shall argue that skilled actions satisfy the definition of basic action; hence the class of basic actions is larger than that of mere bodily movements. I shall also argue that our understanding of action is not furthered by taking either action or basic action to be a primitive term. I shall analyze overt or physical action as volition plus physical change.

Annette Baier, in "The Search for Basic Actions,"<sup>2</sup> attacks the concept of basic action on grounds of incoherence. Most relevant to my present purpose is her example of "the gestalt lace-tie who can successfully tie his laces, but could not show you the movements by which he did so except by actually retying the laces." Of this example Baier says,

When the gestalt lace-tie is asked to move his hands in a lace-tying way, he must tie his laces. He ties his laces to move his hands, in precisely the sense Chisholm requires. We must presume that the "to" relation is asymmetrical, otherwise it cannot select any actions as basic. We, therefore, must conclude that the lace-tie does not move his hands to tie his laces. The lace-tying, then, is basic (as long as there is nothing else

he does to do it), and the hand-moving is non-basic, thus violating A2.<sup>3</sup>

A2 is the requirement that a criterion for basic action select only bodily movements as basic.

Baier uses this example to illustrate the incoherence of basic action theory; the definition of basic action conflicts with the assumption that only bodily movements are basic. The concept of basic action can be maintained, however, by discarding the assumption which Baier codifies as A2. We are then free to assert that the gestalt tie's action is basic. For Baier, the ground for this claim is the asymmetry of the "to" (or "in order to") relation. Lest the reader have his doubts about this asymmetry, I append an additional argument. It begins with the assumption that all actions are intentional. An item of behavior is an action if and only if there is a description under which it is intentional, and only such a description is a description of an action. It follows that a human agent will know such a description of his behavior and, under normal circumstances, be able to state it. Hence if a person manifests certain behavior, *b*, and is unable to give a description of *b*, it follows that *b* is not an action. The sort of inability of which I speak is such as arises from a lack of knowledge of *b*, not from sudden and total paralysis or from the fact that *b* consisted of a successful suicide.

There are many items of human behavior that are constituents of actions but which, on this criterion, do not themselves have the status of action. No cyclist knows or is able to describe just how he moves his body and keeps his machine in balance. Most fluent speakers of English do not know and are unable to describe the movements of the lips and tongue that are the sufficient conditions of speech. Hence, according to my criterion, we

<sup>1</sup> R. M. Chisholm, "Some Puzzles about Agency," in *The Logical Way of Doing Things*, ed. by Karel Lambert (New Haven, 1969), p. 210. See also his "The Descriptive Element in the Concept of Action," *The Journal of Philosophy*, vol. 61 (1964), pp. 613-624, and "Freedom and Action," in *Freedom and Determinism*, ed. by Keith Lehrer (New York, 1966), reprinted in *The Nature of Human Action*, ed. by Myles Brand (Illinois, 1970). See also A. C. Danto, "What We Can Do," *The Journal of Philosophy*, vol. 60 (1963), pp. 435-445, and "Basic Actions," *American Philosophical Quarterly*, vol. 12 (1965), pp. 141-148, and "Causation and Basic Actions," *Inquiry*, vol. 13 (1970), pp. 108-125. "Basic Actions" is reprinted in Brand, *op. cit.*

<sup>2</sup> *American Philosophical Quarterly*, vol. 8 (1971), pp. 161-170.

<sup>3</sup> *Ibid.*, p. 166.

cannot say that the cyclist acts in moving his body in a certain manner in keeping the machine in balance, or that a speaker acts in moving his lips and tongue in a certain way in saying "bubble." If my criterion is rejected and things such as these *are* said, an important distinction is slurred over. Thus cycling and speaking are basic actions. By the same criterion, the fact that the gestalt tier cannot describe his manual operations shows that his act of trying is basic. But it was not always basic. In this respect his case differs from that of the cyclist and the speaker.

It is my view that as skills are acquired, actions that are non-basic at the beginning of the learning process become basic; thus a person's repertoire of basic actions is increased and enriched. "Increased" here refers to the number of different basic actions in the repertoire—lace-tying, typing, piano-playing, etc. "Enriched" refers to growth in the degree of expertise in a given skill.

As an example of the change of an action from non-basic to basic, think of a child learning to tie a shoelace. He is shown that he must (i) grasp the ends, one with each hand; (ii) cross the two end parts; (iii) pass the first end-part under the second; (iv) pull the two ends tight; (v) form a loop in the first end-part; (vi) pass the second end-part fairly loosely around the loop; (vii) make a loop in the second end-part; (viii) pass this loop under the part of the second end-part that is around the first loop, and, finally, (ix) pull the two loops tight. On this account he learns to perform a sequence of nine actions. But there is nothing sacrosanct about the number nine. Suppose the child is confused about step (v), the forming of the first loop. His instructor might say, "(a) Grasp the bottom of the first end-part with the thumb and finger of your right hand, like *this*; (b) run your thumb and finger about an inch up the lace, like *this*; (c) grasp the lace with your left hand about an inch further up the lace, like *this*; (d) move the piece that is in your left hand to the bottom of the lace, like *this*; and (e) take hold of the other part of the lace as well." What was originally described as one action has been expanded to a sequence of five! How many actions does it take to tie a lace? It depends, in a way that I shall try to make clear, upon the person doing the tying. In any case the child's first action of tying his laces is non-basic. There are actions (just how many we cannot say) that he performs in order to tie them, and by performing these actions he causes the laces to be tied.

As time passes, the child becomes increasingly

skillful in tying his laces until, let us suppose, he becomes a gestalt lace-tier. I shall sketch how such a thing might happen. It is a matter of common experience that the acquisition of a skill is often accompanied by a forgetting of the details of how the skilled action is performed. In order to teach my son, John, to operate the standard shift in our car, I found that I must "watch myself" shifting and analyze my technique so that I could describe it. I was wrong in my first account of how I engage the clutch; John and I suffered through jerky starts until I saw my mistake and told him the correct story. When I learned to drive I followed detailed instructions; as my skill increased the details were forgotten and I came to engage the clutch as a basic action. Similarly, the action of tying a lace was non-basic for the child; but it is basic for the gestalt lace-tier that he becomes. The manual behavior that was originally a sequence of actions performed in order to tie the laces has lost the status of action—with the possible exception of step (i). Given that he has grasped the ends, there is nothing further that the gestalt tier does in order to tie the lace. He simply ties it.

This ultimate sort of expertise would come into being gradually. We have seen that although lace-tying, for the beginner, is a sequence of actions, we cannot specify just what these actions are or how many members are in the sequence. Is it one action or five to form a loop? A beginner who learned a sequence of five might, in time, become a gestalt loop-former, unable to tell us how to form a loop except by actually forming one. We might suppose, further, that the forming of the loop (step (v)) would become continuous with the passing of the other end around the loop (step (vi)), so that these two "steps" would cease to be distinguished by the agent as constituting two distinct actions. This process of collation of "steps" would continue until the entire tying operation became one continuous action.

My lace-tying and clutch-releasing examples might suggest that a skilled action is basic only if the agent is unable to describe the details which, for the beginner, are its constituent actions. I have used such cases as examples in order to enhance the persuasiveness of my argument; but I do not intend the consequence that my own lace-tying, which is just as skillful as that of the gestalt tier, is nevertheless non-basic action because I can describe how I do it. I am, indeed, able to rehearse a relevant sequence of steps, but not in the approximately two seconds that I spend in tying a lace. The



gestalt tier and I are alike in performing one continuous action; we differ in that I can analyze this action into steps and he cannot.

I conclude that it is possible and useful to distinguish between basic actions—those things which we do without having to do other things to get them done—and non-basic actions. The class of basic actions, however, is not restricted to mere bodily movements; it includes skilled actions. Hence, although persons in a given cultural setting will have repertoires of basic actions that have much in common, it is unlikely that any two persons will have precisely the same repertoire. What is the usefulness of this distinction? It is a conceptual stepping-stone; the arguments of this section are a helpful preliminary to my theory of volitions.

Since writing the preceding I have read Jane Martin's recent paper, "Basic Actions and Simple Actions."<sup>4</sup> This author, troubled about the same points that bother me in basic action theory, distinguishes between actions and "doings"; the latter are bodily movements which are components of actions but which, themselves, lack the status of action. This is an important distinction; I hope that my theory of volitions helps us to understand its significance.

## II. VOLITIONS

What is the difference between action and mere behavior, between moving one's arm and one's arm moving? I shall argue that the difference is best analyzed by saying that *volition* is present in the former case and absent in the latter. On this assumption of mine, action can be analyzed into two components, volition and physical change. I shall begin with a definition of volition which limits the application of this concept to the context of physical action. *A volition is a conscious or mental act which initiates and guides the physical change that is brought about deliberately in a physical act.* This volition is a component of the physical action. One action can, of course, be a component of another. The child's act of forming a loop is a component of his act of tying his lace; a beginner-typist's act of depressing the 'b' key is a component of his act of typing the letter 'b'. One consequence of this definition is that Wittgenstein's question, "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"<sup>5</sup> is senseless. The fact that

the typist depresses the 'b' key is not to be subtracted from the fact that he types 'b'.

In calling volition a mental act, I do not presuppose dualism; I use the mental-physical distinction simply to mark out the obvious difference between "private" actions, such as planning a bank-robbery, and "public" ones such as robbing a bank. My position is so far from being dualistic that I suggest below that volition be identified with a neurophysiological process. I make this suggestion solely for reasons of simplicity, not because of any commitment to materialism.

To lend some plausibility to my claim, I shall begin by giving examples of the presence of volition in human experience. The first is the case of a bright, spastic child named Valerie who was once a member of my wife's Brownie pack. My wife tried to teach the child to skip. One day Valerie exclaimed in despair, "Mrs. Ripley, I think I'll never learn to skip! When I tell my legs to jump, I can't tell my arms to turn the rope, and when I tell my arms to turn the rope, I can't tell my legs to jump!" The girl's pessimism was justified; she did not succeed in learning to do both things at once. Her handicap made it impossible for her to make the appropriate movements without "telling" her limbs to make them, and the "telling" required a concentration of attention that could be given only to one set of limbs at a time.

Valerie's "telling" is, on my interpretation, her description of volition. Her metaphor is, unfortunately, misleading. An officer tells a subordinate to perform a task; the subordinate, not the officer, performs it. But Valerie does not issue a message to her legs and then leave them to their own devices; she not only initiates their motion but also guides it. The volition, "telling her legs to jump," occurs over the same space of time as the jumping. She jumps in a certain manner, guides her feet over the rope, takes care to land in such a way that she can jump again immediately, and so on. Hence I speak of the volition as a component, not an antecedent, of the physical action.

Experiences such as Valerie's are not restricted to handicapped persons. In an attempt to improve my swimming I am presently learning the front crawl. Part of this stroke is the flutter kick, done from the hips with knees stiff. I find that to do the kick properly I must deliberately hold my knees stiff, or "tell them to remain stiff." If I turn my attention

<sup>4</sup> *American Philosophical Quarterly*, vol. 9 (1972), pp. 59-68.

<sup>5</sup> *Philosophical Investigations*, vol. I, p. 621.

to my arm stroke, observers tell me that my knees are bending. But my case differs from Valerie's in that I can look forward to the time when I do a proper flutter kick without "telling my knees to remain stiff." When a normal person swims or skips, he cannot attend specifically to all of the movements of which his complex behavior is composed. Expert swimming or skipping is a basic action. We shall see in due course what is the role of volition when an action that is originally non-basic becomes basic.

Volition is also present in familiar, everyday actions such as moving one's arm. This claim is made plausible by two examples presented by G. N. A. Vesey in his paper, "Volition."<sup>6</sup> The first, taken from William James, has to do with the effects of anaesthesia.

William James reports that if a patient who has lost sensation in one arm is asked to put the affected hand on top of his head while his eyes are closed, and is at the same time prevented from doing so, he will be very surprised, on opening his eyes, to find that the movement has not taken place. This, James says, is the rule rather than the exception in anaesthetic cases. Confirmation of this is to be found in case-histories reported in the medical journals.<sup>7</sup>

Why did the individual think he had placed his hand on his head when, as a matter of fact, he had not? It would seem that, to put the matter in Humian language, he had the internal impression that he normally has when he knowingly places his hand on his head. Or, in my terminology, he performed the conscious or mental act that normally initiates and guides the motion of his hand to the top of his head. Some such analysis is suggested by Vesey's final description of this case; he answers Wittgenstein's question by saying,

So far as he, but not necessarily his arm, was concerned, he moved his arm.<sup>8</sup>

His second example concerns amputees who are able to "move" their "phantom limbs."

When medical people talk of a "phantom limb" they are using the word "phantom" in a technical sense to refer to the seeming persistence of a part that was known to be gone, and the continued possession of

many of its original attributes of form, position, and even voluntary movement.<sup>9</sup>

A person who has lost an arm retains a "sense" that his arm is present and movable.

Attempts to move painful phantom limbs usually result in an increase of the pain, but voluntary movement of painless phantom limbs 'is usually possible and often considerable.'<sup>10</sup>

The amputee who "moves" a "phantom limb" is, in my terminology, performing a conscious or mental act which, prior to the amputation of the limb, initiated and guided its movement.

The examples borrowed from Vesey differ from my earlier examples in that it seems inappropriate to use locutions such as "telling my arm to move," or "telling my hand to move to the top of my head." Ordinary arm-movements are accomplished with ease. The mental effort they require is so minimal that it escapes the notice of the normal agent; in order to discover its presence we must have recourse to abnormal cases. Having discovered it there, I assume it to be present also in the act of a non-anaesthetized person who puts his hand on top of his head, and in that of a non-amputee who moves his flesh-and-blood limb. Thus it is my claim that volition is a component of all physical actions; they are appropriately analyzed into two components, volition and physical change.

There are two principal and well-founded objections to the notion of will or volition. Traditionally, a volition is a mental act which is a causal antecedent of the relevant physical change; the volition and its consequent are distinct events. Ryle argues effectively that there are no such mental events in our experience; for there is no use for the expressions that we should expect to characterize them if they were present.<sup>11</sup> My view differs from the traditional one in regarding volition as a component of the relevant physical act rather than as its antecedent. Normally there is no volition distinct from the physical act, any more than there is a key-depressing act that is distinct from the typing act. Thus far I agree with Ryle; but I reject his conclusion that the concept of volition is therefore to be discarded.

The notion of volition is rejected also because of

<sup>6</sup> G. N. A. Vesey, *Essays in Philosophical Psychology*, ed. by D. F. Gustafson (New York, 1964), pp. 41-57.

<sup>7</sup> *Ibid.*, p. 41.

<sup>8</sup> *Ibid.*, p. 55.

<sup>9</sup> *Ibid.*, p. 42.

<sup>10</sup> *Ibid.*

<sup>11</sup> Cf. G. Ryle, *The Concept of Mind* (New York, 1949), pp. 64-65.

its traditional affinities with dualism. Vesey, thinking that the notion of willing is inextricably bound up with dualism, rejects the proposition, "I willed the movement to occur," as an answer to Wittgenstein's question.

... the mental state of willing consists in having an idea of the movement desired, which somehow triggers off the causal chain ending in the movement (the so-called 'ideo-motor theory') ...<sup>12</sup>

On this basis he argues persuasively that

... what lies behind the use of the word 'will' is the thought that the mind is only accidentally embodied.<sup>13</sup>

Vesey believes, and I agree, that this "thought" does not do justice to our experience.

The correctness of Vesey's diagnosis of this malady of traditional volitional theory is illustrated by the following passage from Prichard.

There is no reason to limit the change which it is possible to will to a movement of some part of our body since, as James says in effect, we can just as much will the sliding of a table towards us as a movement of our hand towards our head. ... And it looks as though we sometimes will such things in ordinary life, as when in watching a football match we want some player's speed to increase, and will it to increase.<sup>14</sup>

Prichard regards the "willing" involved in moving my hand as being the same sort of mental activity as "willing" that the table slide toward me. In this sense I might also "will" that my ears flap like wings. I think it is a major error to assimilate the volition involved in normal hand motion to "willing" in this sense. One decisive difference is that, as Danto would say, I do not know what it would be like to flap my ears.<sup>15</sup> Suppose that I ask Valerie to "tell" her legs to jump. She jumps. Then I ask her to "tell" her ears to flap. In puzzlement she replies, "But I don't know how." An ear-flapping volition would be a conscious initiation and guidance of the motion of one's ears. Since a volition is a component of an overt action, only a person able to flap his ears could exert an ear-flapping volition. But anyone can "will" that his ears flap—or, for that matter, that his neighbor's ears flap. In neither case, however, are the ears likely to flap. Whereas volition normally effects the physical change that is its object; such is not the case with "willing."

Although I regard the physical changes effected deliberately in overt actions as being brought about by volition, this "bringing about" relation is difficult to describe in terms of cause and effect. The problem is that we can normally state a sufficient condition for the physical change, a condition that is not a volition and not identifiable with a volition.

Let us think of a beginner-typist typing the letter 'b'. Observing his bodily movements, we see that the second finger of his left hand moves to the 'b' key and depresses it; a hammer then strikes the paper and prints 'b'. The motion of the finger causes the letter to be printed. Presumably the finger-motion is caused by a muscular contraction and the contraction by a neural event. We think of a causal chain that begins in the agent's brain and culminates in the observable movements I have described. The consideration of present importance is that *the muscular contraction is a sufficient condition for the motion of the finger*. But we can also say of this typist that he moves his finger to the 'b' key—that he initiates the motion of his finger and deliberately guides it to that place. This I take to be a fact of experience which can be described by saying that *it is the typist's volition that brings about the motion of the finger to the key*. The volition and the consequent finger-motion are experienced as being related immediately; in other words the finger motion is a basic action.

Now, how are these accounts to be made consistent? The volition "brings about" the motion of the finger; the muscular contraction causes it. It will not do to identify the volition with the contraction; for we do not wish to hold that the volition is caused by a neural event in the hand. Moreover, the example of the amputee shows us that a volition can occur when the relevant muscles do not exist.

Nor will it do to identify the volition with the first event in the causal chain. We would have no way of accounting either for the apparent immediacy of the relation between volition and bodily movement or for the fact that *the object of volition changes as the skill of the agent increases*. I shall now direct the reader's attention to this latter important fact. By the object of volition I mean the physical change that is deliberately and consciously initiated and guided by the agent in the course of the rele-

<sup>12</sup> Vesey, *op. cit.*, p. 24.

<sup>13</sup> *Ibid.*, p. 55.

<sup>14</sup> H. A. Prichard, "Acting, Willing, Desiring," in Brand, *op. cit.*, p. 46.

<sup>15</sup> Cf. A. C. Danto, "Basic Actions," in Brand, *op. cit.*, p. 267.

vant action. In the case of the typist, the object of volition is the motion of his finger.

Let us turn from the beginner to an expert whose typewriter chatters like a machine-gun as he types. He types the title of an essay, "Philosophic Rubbish in Bacon and Hobbes." In doing so he types the letter 'b' five times. In each case the motion of the finger that types 'b', like the loop-forming of the gestalt lace-tie, is not an action. I support this claim by an argument similar to that used in the lace-tying case. It is my own experience that when I am confronted with a typewriter with unmarked keys and asked to tell which is the 'b' key, which the 'n' key, etc., I am unable to do so, even though I am able to type. I must type or simulate the typing of a word containing 'b' in order to "catch myself in the act" and identify the 'b' key by watching my finger depress it. For the beginner it is an action to guide his finger to a certain key and depress it, as a means of typing 'b'; an accomplished typist performs no such action. Even if the expert is a typing teacher who knows well which keys are which, he cannot deliberately guide his finger to the 'b' key and depress it when typing at such a speed that his typewriter chatters like a machine-gun.

My position draws a sharp contrast between the beginner and the expert; I argue that as they type they are not performing the same actions. Current action theory, with its behavioristic orientation, would hold that their actions are the same, and that the observable differences in their behavior are to be described adverbially. In support of my view I would point out how striking are these observable differences. The beginner types slowly, hesitantly, haltingly, and with concentrated effort. For all of his slowness and his exertion of care and effort, he types inaccurately. The expert, on the other hand, types with machine-gun rapidity, confidently, rhythmically, and with ease. For all of his break-neck speed and relaxed demeanor, his work is accurate. I think that these differences are sufficiently great to lend a high degree of plausibility to the contention that these two typists are performing different actions. To say that they are performing the same action in different ways is to make it difficult to draw important distinctions.

Let us describe the process of learning to type using the concept of volition. Let us think of the act of typing "rubbish." For the beginner, each object of volition is a movement of a certain finger to a certain key and the depressing of the key by the finger. He starts out by initiating and guiding the

motion of his first finger, left hand, to the 'r' key, and depressing this key. Then he initiates and guides the motion of his first finger, right hand, to the 'u' key and depresses it. And so on. For the pupil of intermediate skill, the objects of volition are (i) the typing of 'r', (ii) the typing of 'u', etc. For the machine-gun typist the relevant object is the typing of "rubbish" or, perhaps, of the whole sentence in which "rubbish" occurs. In general, the intermediate pupil performs no action of guiding his finger to the appropriate key; the expert performs no action of typing a single letter. As a typist's proficiency increases, there is a change in the actions he performs and hence in his volitions. To put the matter more specifically, the object of his volition changes from (a) the motion of a finger in order to type a letter that is part of a word to (b) the typing of a letter that is part of a word, and from (b) to (c) the typing of a word. We might say that as the agent's skill increases the object of volition moves "outwards," from cause to effect and from part to whole.

All the while, however, the causal sequence remains the same. A causal account of the expert's behavior in typing 'b' is exactly the same as that which was given in the case of the beginner. In both cases we think of a causal chain beginning with an event in the brain and culminating in the contraction of muscles, the motion of a finger, the depressing of a key, and the motion of the hammer that prints 'b'. If the volition is identified with the first event in this chain—or, for that matter, with any member of it whatsoever—we are powerless to give an account of the change in the object of volition that occurs as I have just described.

The upshot of this discussion is that the notion of a causal chain, a sequence of discrete events which links up a brain event with an item of overt behavior, is not helpful for the elucidation of our experience. The assumption that there is such a chain has given rise to conceptual puzzles that seem to be insoluble. In this paper I can do little more than to make a bare suggestion of an alternative line of thought. My suggestion is that the puzzles dissolve if we think in a *holistic manner* of the process linking brain event with behavior. The volition can then be identified, not with a "link" in a supposed "causal chain," but with the entire process that begins in the brain and effects the object of volition. On this view the muscular contraction, in the case of both typists, is a *component* of his volition. There is no longer a conflict between the well-grounded assertions (i) that the muscular

contraction is sufficient for the finger-motion, and (ii) that the finger-motion is brought about by volition. It is also possible to make sense of the changes that are observed to occur in the object of volition as skill increases. In the case of the beginner, the moving of the finger to the 'b' key is an *object* of volition; but in the case of the expert it is a *component* of his volition rather than its object. The volition, regarded as a holistic process, can plausibly be thought to extend itself or "grow" by taking in or assimilating its object. Is this an unacceptably odd way of speaking? I think not. We do talk in this manner of cells, organs, and complex organisms. Why should such language be inappropriate when applied to a neurophysiological process?

In identifying volition with a neurophysiological process, I am influenced by the non-reductive materialism sketched by Richard Taylor in his "How to Bury the Mind-Body Problem."<sup>16</sup> I recognize that such an identification involves conceptual problems; but I think they are less formidable than those which beset alternative theories. The possible alternatives to my view are non-

volitional theories and dualistic volitional theories. With regard to the former, I have argued that we need the concept of volition to make rational sense of our experiences of various kinds of action. With regard to dualism, I shall neither repeat any part of the vast corpus of antidualistic arguments nor attempt to add to it. It is an open question whether or not it is possible to formulate a coherent, dualistic account of the data described in this paper. Since my bias is antimaterialistic I have tried to formulate one, but have been frustrated at every turn by conceptual tangles. Suffice it to say that my suggestion that volition be identified with a neurophysiological process, holistically conceived, gives rise to a theory that is much less complicated than a dualistic theory would have to be—if a dualistic theory is indeed possible. Until there arises a compelling need to postulate a mind-substance, I shall, for reasons of parsimony, refrain from doing so.

*Lakehead University*

*Received February 27, 1973*

<sup>16</sup> *American Philosophical Quarterly*, vol. 6 (1969), pp. 126-143.

## BOOKS RECEIVED

- Arndt, Hans Werner, *Methodo Scientifica Pertractatum* (Berlin and New York: Walter de Gruyter, 1971), 170 pp.
- Brody, Baruch A., *Logic: Theoretical and Applied* (Englewood Cliffs: Prentice-Hall, Inc., 1973), 280 pp.
- Bunge, Mario, *Philosophy of Physics* (Dordrecht: D. Reidel Publishing Co., 1973), 248 pp.
- , *Method, Model and Matter* (Dordrecht: D. Reidel Publishing Co., 1972), 196 pp.
- Cesa, Claudio, *Studi Sulla Sinistra Hegeliana* (Urbino: Argalia editore urbino, 1972), 365 pp.
- Hamesse, Jacqueline, *S. Bonaventure* (Louvain: Université Catholique de Louvain, 1972), 193 pp.
- , *Auctoritates Aristotelis* (Louvain: Université Catholique de Louvain, 1972), 230 pp.
- Illuminati, Augusto, *Società e Progresso Nell'illuminismo Francese* (Urbino: Argalia Editore Urbino, 1972), 265 pp.
- Kaufmann, Walter, *Without Guilt and Justice* (New York: Peter H. Wyden, Inc., 1973), 294 pp.
- King-Farlow, John, and Christensen, William N., *Faith and the Life of Reason* (Dordrecht: D. Reidel Publishing Co., 1972), 253 pp.
- Kurtz, Paul, and Dondeyne, Albert (ed's), *A Catholic/Humanist Dialogue: Humanists and Roman Catholics in a Common World* (New York: Prometheus Books, 1973), 117 pp.
- Kurtz, Paul (ed.), *Moral Problems in Contemporary Society* (New York: Prometheus Books, 1973), 301 pp.
- Lines, John Davis, *The Duality of Physical Truth and Cause* (New York: Philosophical Library, 1973), 94 pp.
- Martin, Gottfried, *Arithmetik und Kombinatorik bei Kant* (New York and Berlin: Walter de Gruyter, 1972), 165 pp.
- Olafson, Frederick A., *Ethics and Twentieth Century Thought* (Englewood Cliffs: Prentice-Hall, Inc., 1973), 134 pp.
- Pande, G. C., *The Meaning and Process of Culture* (Agra: Shiva Lal Agarwala & Co., 1972), 179 pp.
- Reiter, Josef, *System und Praxis* (München: Verlag Karl Alber Freiburg, 1972), 275 pp.
- Robinet, André, *Le Defi Cybernetique* (Gallimard: Editions Gallimard, 1973), 232 pp.
- Rotenstreich, Nathan, *Philosophy: The Concept and Its Manifestations* (Dordrecht: D. Reidel Publishing Co., 1972), 258 pp.
- Solomon, Robert (ed.), *Nietzsche* (New York: Doubleday Anchor, 1973), 391 pp.
- Tartaglia, Philip, *Problems in the Construction of a Theory of Natural Language* (Paris: Mouton, 1972), 252 pp.
- Tondl, Ladislav, *Scientific Procedures* (Dordrecht: D. Reidel Publishing Co., 1973), 268 pp.
- Wallace, William A., *Causality and Scientific Explanation* (Ann Arbor: The University of Michigan Press, 1972), 288 pp.
- Zeman, J. Jay, *Modal Logic* (Oxford: The Clarendon Press, 1973), 302 pp.

# The Personalist

John Hospers, *Editor*

AN INTERNATIONAL JOURNAL OF PHILOSOPHY

Contents of Autumn, 1973, issue (Volume 54, No. 4)

## Ethics Issue

### Articles

- |  |                               |
|--|-------------------------------|
| I. The Complexity of 'Why Be Moral?'                         | <b>Michael D. Bayles</b>      |
| II. How to Establish or Refute Ethical Relativism            | <b>Thomas McClintock</b>      |
| III. The Pursuit of Happiness                                | <b>Richmond Campbell</b>      |
| IV. No Percepts of Morality                                  | <b>Michael H. Robins</b>      |
| V. Is Population Control a Difficulty for the Utilitarian?   | <b>Milton Goldinger</b>       |
| VI. Moore's Defense of the Rule 'Do No Murder'               | <b>Ronald Suter</b>           |
| VII. Kant: Two Concepts of Moral Ends                        | <b>Fred D. Miller</b>         |
| VIII. Rule-Egoism  | <b>John Hospers</b>           |
| IX. The Institutional Theory of Obligation: A Brief Critique | <b>Sheldon P. Peterfreund</b> |

Books Received

Our Contributors' Page

Notices

**Subscriptions** \$6.00 per year, \$2.00 per issue

Send to Mrs. Daphne Garey, Managing Editor, School of Philosophy, University of Southern California,  
University Park, Los Angeles, California 90007, USA

## NOUS

Nihil philosophici a nobis alienum putamus

Volume VII

Number 4

November 1973

Richard Rorty / *Criteria and Necessity*

Nicholas Rescher and Zane Parks / *Possible Individuals, Trans-World Identity, and Quantified Modal Logic*

Baruch Brody / *Why Settle for Anything Less Than Good Old-Fashioned Aristotelian Essentialism*

Kenneth Friedman / *Son of Grue: Simplicity vs. Entrenchment*

Kathryn Parsons / *Ambiguity and the Truth Definition*

Adam Morton / *The Possible In the Actual*

Lou Goble / *Opacity and the Ought-to-Be*

Recent Publications in Philosophy

Index to Vol. VII

# INQUIRY

An Interdisciplinary Journal of Philosophy and the Social Sciences

Edited by Arne Naess and Alastair Hannay

Number 4

Winter 1973

Volume 16

On What it is Like to be a Man

Lao Tzu's Conception of Tao

Rationality and the *Tu Quoque* Argument

Knowledge *v.* Survival

Discussions:

Comments on 'Knowledge *v.* Survival'

Rejoinder to Naess

Cody's Categories

Reply to Mr. Døør

Virgil C. Aldrich

Charles Wei-hsun Fu

Joseph Agassi

Herman Tennessen

Arne Naess

Herman Tennessen

David Rayfield

Peter Zinkernagel

Review Discussions:

Wittgenstein Reinvestigated (James Bogen: *Wittgenstein's Philosophy of Language: Some Aspects of its Development*; P. M. S. Hacker: *Insight and Illusion: Wittgenstein on Philosophy and the Metaphysics of Experience*)

Understanding Hermeneutics (Richard E. Palmer: *Hermeneutics: Interpretation. Theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*)

John E. Llewelyn

Guttorm Fløistad

Published quarterly by Universitetsforlaget, P.O. Box 307, Blindern, Oslo 3, Norway

Subscription (incl. postage): N. kr. 75, \$12.50, £4.90 (Institutions)

N. kr. 60, \$10.00, £3.90 (Individuals)

## The Philosophical Quarterly

CONTENTS OF VOLUME 24

NO. 94

JANUARY 1974

*Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant and Stewart McNeilly.*

*Published for the Scots Philosophical Club by the University of St Andrews*

### ARTICLES

I FOR EVER?

R. F. Holland

II CATEGORIES IN NATURAL LANGUAGE

A. G. Elgood

III LOGIC OF TERMS

G. D. Duthie

IV REASONS FOR ACTION

Scott Meikle

V THE TRUTH ABOUT

FICTIONAL ENTITIES H. Gene Blocker

### DISCUSSIONS

MILL'S ACT-UTILITARIANISM

D. G. Brown

ON THE "GEACH-MOUTON LINE"

David L. Mouton

### BOOK REVIEWS

### LIST OF BOOKS RECEIVED

Annual Subscriptions: Individuals £3.00 (\$8.50) Institutions £5.00 (\$14.00)

Single Issues £1.00 or \$3.00 (Institutions £1.50 or \$4.00)

for orders placed directly with the Journals Manager, Scottish Academic Press

Ltd., 25 Perth St., Edinburgh EH3 5DW, Scotland



# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich  
Annette Baier  
Kurt Baier  
Stephen F. Barker  
Monroe Beardsley  
Nuel D. Belnap, Jr.  
Myles Brand  
Robert Butts  
Roderick M. Chisholm  
L. Jonathan Cohen

James Collins  
Antonio S. Cua  
Arthur C. Danto  
Richard M. Gale  
Peter Thomas Geach  
Marjorie Grene  
Adolf Grünbaum  
Rom Harré  
Carl G. Hempel

Hugues Leblanc  
Keith Lehrer  
Gerald J. Massey  
George Mavrodes  
Richard Rorty  
George A. Schrader  
Wilfrid Sellars  
Ernest Sosa  
Richard Taylor  
John W. Yolton

---

VOLUME 11/NUMBER 3

JULY, 1974

## CONTENTS

- |   |     |   |     |
|---|-----|---|-----|
| I. WILLIAM A. PARENT: <i>Some Recent Work on the Concept of Liberty</i> . . . . .             | 149 | V. JOYCE TREBILCOT: <i>Aprudentialism</i> . . . . .                             | 203 |
| II. DON LOCKE: <i>Reasons, Wants and Causes</i> . . . . .                                     | 169 | VI. STEPHEN SCOTT: <i>Rituals and "Ought"</i> . . . . .                         | 211 |
| III. J. MICHAEL YOUNG: <i>The Ontological Argument and the Concept of Substance</i> . . . . . | 181 | VII. JOHN FISHER: <i>Universalizability and Judgments of Taste</i> . . . . .    | 219 |
| IV. ROBERT YOUNG: <i>Petitioning God</i> . . . . .  | 193 | VIII. RUTH BARGAN MARCUS: <i>Classes, Collections and Individuals</i> . . . . . | 227 |

# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$14.00, increasing by \$1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$11.20, increasing by \$1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of \$6 per group of 50.

## SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers (including the annual Monograph) and eighteen dollars for institutions (excluding the annual Monograph which is obtainable on payment of another seven dollars). Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and five dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

## MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



## I. SOME RECENT WORK ON THE CONCEPT OF LIBERTY

WILLIAM A. PARENT

IN this paper I want to examine recent work (1958 to the present) on the concept of social and political freedom, or liberty. Since the publication almost 15 years ago of *Two Concepts of Liberty*, Sir Isaiah Berlin's Inaugural Address to Oxford University, many articles and books have been written with the principal aim of elucidating the nature of liberty. Because Berlin's ideas have generated so much of the contemporary interest in this subject and have had such an impact on the approach and thinking of other philosophers concerned with the problem, I shall begin the survey with a detailed evaluation of them. Unfortunately Berlin's essay, as well as most of the other recent writings to be discussed, fail miserably in their common attempt to shed light on the nature of liberty; instead they are replete with all sorts of conceptual muddles and inconsistencies. In the following analyses I will point out these confusions and inconsistencies and try to indicate some of their possible sources. My own views regarding the meaning of liberty will become apparent as the survey proceeds.

### I. BERLIN'S TWO CONCEPTS OF LIBERTY

In his Inaugural Essay Berlin discusses the two concepts of liberty which, in his opinion, stand out as centrally important among the more than two hundred different senses of the term recorded by historians of thought. He calls these the negative and positive concepts. I shall first summarize Berlin's discussion of them, beginning with the former, and then will present my objections to his principal contentions.

Negative freedom is involved in answering the question of the area within which persons should be left to do what they want without interference by others. To be negatively free essentially means not to be interfered with in the pursuit of one's desires. In Berlin's words: "I am normally said to be free to the degree to which no human being

interferes with my activity. Political liberty in this sense is simply the area within which a man can do what he wants" (7).<sup>1</sup> Hence anyone who deliberately prevents me from satisfying my desires, attaining my goals (7), fulfilling my wishes (8), or making choices (16)—Berlin uses these phrases interchangeably—infringes upon my freedom.

That such interferences must be deliberate is implied, Berlin thinks, by the concepts of coercion and enslavement contained in the meaning of unfreedom. Thus if I accidentally lock *A* in his room and thereby make it impossible for him to do what he wants, no one would accuse me of rendering him unfree to leave. But we would, on the other hand, describe a man whose government deliberately makes it impossible for him to buy food as unfree to do so; for he is, while *A* was not, a victim of an insidious form of coercion (8).

Berlin also emphasizes that the mere incapacity to perform some action does not imply unfreedom to perform it. We would not, nor should we, regard the physically handicapped as unfree to engage in strenuous exercise or the mentally retarded as unfree to succeed in college unless it could be shown that their inability to do these things resulted from the deliberate actions of other persons or groups. Berlin's point is well taken, but I shall shortly show that he is wrong in thinking that an inability to act caused by the deliberate activity of others is sufficient for unfreedom.

The second, or positive, concept of freedom Berlin discusses is involved in answering the question what, or who, is or should be the source of control or interference that can determine someone to do or be one thing rather than another. Freedom in this sense consists essentially of rational self-determination or self-mastery. It derives from man's desire to be in control of his own destiny. A proponent of positive liberty wants his life and decisions to depend on himself, not on external forces. He wants to be an instrument of

<sup>1</sup> All page references to works discussed will be placed in parentheses. Information regarding publishers and dates of publication can be found in the bibliography at the end of the essay.

his own, not of others', acts of will; to be a subject, not an object; to be moved by purposes which originate with him, not with someone else.

The concept of self-mastery seems at first glance quite straight forward and innocuous. We understand, or at least think we understand, what it means: we can, for example, point out the kind of people who are obviously not masters of themselves—e.g., drug addicts, alcoholics, and compulsive gamblers. But, as Berlin points out, the concept does naturally suggest and many philosophers have interpreted it to imply a peculiar dualistic theory of the person according to which each of us is composed, on the one hand, of a "real" self, the transcendent, dominant controller, and on the other of a bundle of feelings and passions to be disciplined and brought to heel. Philosophers subscribing to this metaphysical interpretation of rational self-mastery usually identify the "real" self with reason, with man's "higher" or autonomous nature and contrast it with his "lower" animal nature which, on their view, is made up entirely of irrational impulses and uncontrollable desires. Moreover, they are inclined to argue that rational self-mastery, understood to imply this kind of "dualism," constitutes the only genuine purpose of man and therefore should take precedence over and be pursued at the expense of other goals which he might mistakenly think are equally important. Indeed, man can and should be forced to be free in this positive sense, however violently his poor, unreflecting, desire-ridden self may cry out against this process. In Berlin's words: "Freedom is not freedom to do what is irrational, or stupid, or bad. To force empirical selves into the right pattern is no tyranny, but liberation" (32-33).

To this theory of positive freedom Berlin raises two major objections. First, rational self-mastery is not the sole nor even necessarily the most important goal in life. Instead it is just one among several possible goals not all of which are commensurable (56). Secondly, the contention that force utilized to achieve rational self-mastery can be justified, that anyone who does not realize what is truly good for him should be made to realize it, is morally bankrupt. To oppress, torture, or coerce men in the name of their freedom, far from being justifiable, in fact betrays the most dangerous kind of moral despotism. For Berlin, power can never rightfully be exercised over any mentally sane adult except to prevent him from harming others (35).

There are, I believe, a number of serious diffi-

culties with Berlin's account of negative and positive freedom as summarized above which have gone largely unnoticed by his students and critics and which unhappily re-appear in many post-1958 writings. They deserve careful consideration here.

For one thing, Berlin's various characterizations of negative unfreedom (the deliberate interference with: any kind of human activity; a person's doing what he wants or wishes; the attainment of his goals; his making choices) are not equivalent. Obviously an activity in which a person is engaged need not be one he either wants or chooses to pursue, nor need it result in the attainment of his goals. Moreover, making choices is not always identical with achieving one's goals (it would only be so for someone whose aim was to be able to choose his own course of action), nor is it something we always want to do. And a man who does attain his goals is not necessarily doing what he wants, in the ordinary sense of this term, particularly if his goals are merely negative in character or pose no real challenge to him. So Berlin introduces at least four different concepts of negative freedom in his essay. Leaving aside the question whether any of them comprises an adequate explication, the point is that he should have followed his own advice not to render the meaning of important terms vague and distended through loose or careless usage (43). As it stands, one cannot always be certain precisely what Berlin means when he refers to *the* concept of negative liberty.

And to confuse matters even more, Berlin occasionally identifies negative liberty, not with any kind of non-interference, but with what results from non-interference, namely an area of personal life entirely immune from outside intrusion. So we have yet another concept of freedom to ponder, a concept which comes very close to expressing the basic idea of privacy. Indeed, Berlin often uses the term "privacy" while discussing negative liberty—e.g., he continually emphasizes the need for private life (10), and equates the feeling we derive from being free with the sense of privacy (11). Clearly, however, liberty and privacy ought not to be equated. Many persons who obviously lack the social freedom to undertake everyday activities, for instance prisoners placed in solitary confinement, nevertheless have plenty of privacy. Berlin's failure to differentiate between these two values is especially puzzling in view of his admonition against confounding related but conceptually distinct political ideals (39).

Only recently has Berlin made it clear<sup>2</sup> that in his Inaugural Address he meant to identify negative liberty with non-interference with a person's desires. In the Address he claims, as if this were a reason for accepting the definition, that all the great English political philosophers conceive of liberty in this way (8). But plainly they do not. Hobbes, for example, equates liberty with the absence of external impediments to motion. This definition is different from Berlin's in at least two respects: it specifies that only a certain class of impediments can jeopardize freedom; and it does not incorporate any reference to a person's doing what he wants—for Hobbes someone might very well be rendered unfree to do what he dislikes.

I believe there are at least two good reasons why we should reject Berlin's definition of liberty. A prisoner is socially unfree to assume the responsibilities of an executive whether he desires to assume them or not. Similarly, a man who locks his wife in their home deprives her of the liberty to leave, even though it is conceivable she does not want to leave anyway. Interfering with a person's desires, though sometimes sufficient to curtail his freedom, is not necessary. And Berlin's definition carries the absurd implication that anyone's freedom can be enlarged merely by decreasing the number of his desires. There are, of course, different ways to restore a man's freedom to act, but the extirpation of his desires is not one of them. Conditioning a slave to want only what he has cannot make him a free man.

I have two final comments respecting Berlin's account of negative liberty. First, it is simply not true that a man can be rendered unfree to act only through the deliberate activity of someone else. If *X* closes up his shop, forgetting that *Y* is still inside, and later is asked for an explanation, he should say, "I temporarily deprived *Y* of his freedom, but I didn't do it deliberately." Secondly, Berlin is also mistaken in believing that anyone who deliberately deprives another of the ability to perform an activity curtails his liberty to do so. *X* might deliberately deprive *Y* of the ability to walk by shooting him in the back, but he does not thereby render *Y* socially unfree to walk; rather he deprives him of the physical capacity to do so. The terms "deliberate deprivation of ability" and "interference with freedom" have different

designata. It is important to remember, however, that a person's liberty to do *x* will prove of little value to him if someone deliberately destroys his ability to exercise it.

Turning now to Berlin's treatment of positive liberty, unquestionably the real object of his attacks is not the ordinary, metaphysically naive sense of "rational self-mastery" but the philosophical interpretation of this concept with its bifurcation of the person into an autonomous self on the one hand and an empirical self on the other. We may agree with Berlin that this interpretation is consistent with and has in the past been used to justify injustice and oppression. But cannot the methods of illogic and spurious metaphysics just as easily be employed to distort the concept of negative liberty and produce justifications of tyranny from it? All we have to do is argue for the existence of a "real" self where "true" desires are quite different from and in basic conflict with those arising from our mere animal nature. It would seem that no concept is entirely immune from this kind of distortion; and since the villain of the piece is metaphysical extravagance, why does Berlin insist on obscuring this fact and attacking the concept of rational self-mastery instead?<sup>3</sup>

Moreover, nothing in the ordinary meaning of rational self-mastery precludes the recognition of other equally important ends. Positive liberals who steadfastly refuse to read any sort of metaphysical dualism into their theory need not and probably would not regard freedom as the only genuine human ideal; instead they would ascribe at least as much value to goods like justice, peace, and equality of opportunity since without them the worth of rational self-mastery would be significantly impaired. So Berlin's contention that positive liberalism is inhumane because it denies a pluralism of values is only relevant when the concept of freedom is subjected to philosophical disfigurement.

The point emerging from all this is that Berlin's essay does not provide any reasons why alcoholics, drug addicts, compulsive gamblers, and other persons in bondage to powerful, often irrational impulses, should not be called unfree to pursue most of the activities which men displaying rational self-mastery in the ordinary sense are able to per-

<sup>2</sup> See Berlin's introduction to his *Four Essays on Liberty* (Oxford, 1969).

<sup>3</sup> On this point I am in agreement with Marshall Cohen's criticism of Berlin. See his "Berlin and the Liberal Tradition," *The Philosophical Quarterly*, vol. 10 (1961), pp. 216-228.

form. The idea that freedom consists essentially of exercising discipline and control over one's impulses, in just the way alcoholics and addicts might, does not survive Berlin's criticisms unmarked.

There is another route Berlin could have followed in criticizing the positive concept of freedom. Instead of attempting to show that the philosophical explication of rational self-mastery leads to tyranny and is inconsistent with a pluralism of values, he might simply have asked: Does rational self-mastery, under any interpretation, reflect what we ordinarily mean by liberty? I believe it does not. Consider, for example, the situation of an addict who is driven to steal in order to support his habit. The addict exhibits little or no rational self-mastery, yet he still has the social freedom to engage in illegal activities; indeed this is precisely the problem confronting law enforcement officials, who must attempt to render him at least temporarily unfree to do so. The loss of rational self-mastery may well, then, lead to, but it is not constitutive of, unfreedom as we ordinarily conceive of it. Conversely, some prisoners might somehow manage to acquire an understanding of and control over their base criminal impulses through rehabilitation. These considerations indicate that liberty cannot be adequately elucidated as rational self-mastery.

## II. BERLIN'S LATEST THOUGHTS ON LIBERTY

We have now to examine Berlin's most recent ideas on liberty set forth in the introduction to his *Four Essays on Liberty* (1969). For our purposes the most significant thing about this introduction is Berlin's realization that liberty cannot adequately be defined as the non-interference with a person's doing what he wants. He comes to reject his old definition because it implies that one way to render a man free consists of extinguishing all his desires, and this is clearly absurd; and it encourages the use of repressive techniques for controlling behavior. Curiously enough, Berlin fails to notice an even more obvious defect of this definition which I mentioned above, namely its implication that a person's freedom cannot be abridged so long as he is doing something he dislikes.

Berlin now maintains that freedom properly conceived is the absence of obstacles, resulting from alterable human practices, to possible choices and activities (xxxix). If these obstructions are the intentional result of alterable human practices,

then in Berlin's judgment it is appropriate to speak of oppression; but not all instances of unfreedom need be oppressive in this sense. So Berlin abandons, and wisely so, his earlier insistence that unfreedom must involve the deliberate interference with human activity. With respect to the kinds of social practices posing the most immediate and dangerous threats to negative freedom, Berlin emphasizes these discriminatory economic and educational policies that promote a situation wherein entire groups are shut off from benefits which are then allowed to accumulate exclusively in the hands of the more advantaged citizens. This sort of discrimination builds walls around classes of men and effectively shuts the door to their future development (xlvi-xlviii).

That Berlin's new definition is inadequate becomes evident upon considering the following facts. Illness, ignorance, bias, and hatred often constitute obstacles to a person's acting in various ways. And they often result from alterable human practices—e.g., unsanitary working conditions have been known to cause illness and racist institutions and customs are responsible for much hatred and bias. Yet we do not say that a person who is prevented from pursuing a given activity by illness, ignorance, bias, or hatred resulting from alterable human practices has been rendered unfree in the political sense to do so. Instead we say that persons often can be deprived of the physical ability, the skill or intelligence, the fair-mindedness, and the compassion necessary for them to undertake different tasks.

Discriminatory economic and educational policies, which according to Berlin portend serious consequences for human freedom, in fact do not render their victims unfree to advance materially, but rather deny them the fair opportunity to do so. A person who is refused a job because of his race and who is prevented from attending decent schools because he comes from the "wrong neighborhood" normally would and should express his moral outrage in the language of opportunity, not freedom. He is being denied the *chance* to work and to acquire a satisfactory education. (Of course persons who are not given the opportunity to fulfill their goals will probably find any freedom they have to do so of little value.)

Interestingly enough Berlin on at least two occasions identifies the concepts of liberty and opportunity, first maintaining that liberty is opportunity for action (xlii), then arguing that the problem of how an over-all increase in liberty in particular

circumstances is to be secured and distributed reduces to the question how the maximization of opportunities is in any concrete case to be achieved (xlvi). It seems evident, though, that someone's having the social freedom to do something is not sufficient to warrant the statement that he has the opportunity to do it. A young man, for example, may be socially free to attend a concert but lack the opportunity to do so either because he is too busy or has not been given any tickets. The terms "liberty" and "opportunity" have distinct meanings. Depriving someone of the liberty to act is not the same thing as refusing him the opportunity to do so, and the problem of increasing and distributing freedom cannot be reduced to a question about the maximization of opportunities.

In sum, our final judgment must be that Berlin's endeavor, spanning over a decade, to elucidate the nature of liberty fails. His belief, yet unexamined, that there are two distinct concepts of liberty which can be isolated from one another and analyzed separately, each on its own merits, has also come under attack in recent writings. Among these, Gerald MacCallum's "Negative and Positive Freedom" is perhaps the most widely discussed and influential, so I will now turn to an evaluation of it.

### III. MACCALLUM'S "NEGATIVE AND POSITIVE FREEDOM"

MacCallum challenges the view that it is possible and useful to distinguish between two kinds or concepts of social freedom, negative and positive. He argues that this distinction has never been made sufficiently clear and at any rate can serve only to obfuscate the really critical issues which need examining if the differences dividing philosophers concerned with freedom are to be understood. MacCallum's own view is that the social freedom of some agent or agents is always freedom from some constraint or compulsion on, interference with, or barrier to doing, not doing, becoming or not becoming something. Social freedom is thus always *of* something (an agent or agents), *from* something, *to* do, not do, become or not become something (314). It refers to a triadic relation whose form can be schematized as follows: *X* is (is not) free from *y* to do (not do, become, not become) *z*.

So instead of concentrating on kinds of freedom and asking which concept is the correct or most humane one, à la Berlin, we should, in MacCallum's judgment, interpret all controversies

about freedom as disagreements over the appropriate ranges of one or more of the *x*, *y*, and *z* variables. The fundamental questions in political philosophy center around these disagreements; to present them as involving the meaning of "freedom" defeats any attempt at rewarding, fruitful argument. In MacCallum's words: "It would be far better to insist that the same (triadic) concept of freedom is operating throughout (disputes over freedom), and that the differences, rather than being about what freedom is, are, for example, about what persons are, and about what can count as an obstacle to or interference with the freedom of persons so conceived," (320—the parentheses are mine). We should, for example, look upon the disagreements between, say, Hobbes's theory of freedom and Rousseau's, or Marx's and Mill's, as essentially differences concerning, among other things, what the nature of a person is, and not as disputes over what it means to be free.

But is it possible to translate all genuine statements about the freedom of agents into the *xyz* schema? MacCallum answers that a statement is not intelligible as a statement about the freedom of agents until one is in a position to fill in the elements of this schema. In the end, he argues, discussions concerning the freedom of agents can be meaningful and rationally assessed only after each term of the triadic relation is specified or at least understood. In support of this argument MacCallum considers a number of familiar idioms which do not contain any explicit reference to two of the three variables and shows that each of them can be understood to be about the freedom of agents because each can be translated into the triadic schema. Thus "free beer" is thought to be intelligible as an expression about human freedom because it is understood to refer to beer that people are free from the ordinary restrictions of the market place to drink without paying for it (316); "the property is free from encumbrance" makes sense as a statement about the freedom of agents because it can be interpreted to mean that owners of land are free from certain well-known restrictions to use, enjoy, and dispose of the land as they wish (316); and "freedom from hunger" can also be regarded as an expression about the freedom of agents for we can understand it to signify a state of affairs where people are free from barriers constituted by various specifiable agricultural, economic, and political conditions to get enough food to prevent hunger (318).

MacCallum's thesis can be formulated more

precisely as follows. If a statement is really about the social freedom of agents, then it can be translated into the *xyz* schema, and conversely, if a statement can be translated into this schema, then it is intelligible as or can be understood to be about the social freedom of agents. MacCallum emphasizes that he is interested only in the conditions under which the talk about human freedom is intelligible. This restriction excludes from consideration uses of "free of" and "free from" in statements like "the sky is free of clouds," where "free" simply means "without" or "rid of" (315).

The most serious difficulty with MacCallum's analysis is that there are meaningful statements of the form "*X* is (is not) free from *y* to do *z*" which do not imply any assertions about the social freedom (unfreedom) of agents. MacCallum's thesis that if a statement can be translated into this triadic schema it can be understood to be about human freedom is false. Consider, for example, the following statements, where '*X*' ranges over persons.

- (a) *X* is free from time-consuming duties to take a vacation.
- (b) *X* is not free from laziness to do construction work.
- (c) *X* is free of prejudice to judge the case.
- (d) *X* is free of guilt feelings to resume a normal life.
- (e) *X* is free from poverty to take a world cruise.
- (f) *X* is free of his partial paralysis to swim.

None of these is about social freedom. The first implies only that *X* has the opportunity to take a vacation; (b) tells us that *X* lacks the energy or drive necessary to undertake physical work—laziness is not the kind of impediment that can deprive a person of his social freedom; (c) is only intelligible as a claim about *X*'s fair and impartial attitude—the presence of prejudice implies unjustified partiality, not social unfreedom; (d) says simply that *X* possesses the peace of mind to lead, once again, a normal life; (e) must be understood to mean that *X* has the financial means to go on a world cruise—it does not imply any statement about his social freedom; and (f) implies only that *X* has the physical ability or power to swim—again paralysis is not an impediment properly regarded as a threat to a man's social freedom.

In short, MacCallum's thesis requires us to

hold that assertions which are unmistakably about human opportunity, fair-mindedness, peace of mind, etc., are also about social freedom, and this is plainly false. A person can have both the liberty to act and the peace of mind to enjoy life but there is no contradiction in supposing that he has the latter without the former. Similarly, determining that *X* has the impartiality or the energy, say, to undertake an activity is not by itself sufficient to establish his liberty to do so. And so on.

One of the important points to emerge from the above analysis is that "freedom" in the social sense is not a substantive corresponding to the adjectives "free from" and "free of" as redness corresponds to the adjective "red."<sup>4</sup> Whenever you describe something as red it is always possible to use "redness" also; but "free of" and "free from" are applicable in many cases where "liberty" is not. Whether a statement of the form "*X* is free from *y* to do *z*" is intelligible as a statement about *X*'s social freedom depends upon the referent of the *y* variable, for only certain kinds of preventing conditions are commonly regarded as threats to a person's liberty. Thus we say that a man behind prison bars has been deprived of his liberty to pursue most everyday activities, but we do not say of someone whose hatred stops him from forgiving others that he has been rendered socially unfree to forgive them—what he lacks is the compassion necessary to do so. MacCallum should have carefully discriminated among different kinds of preventing conditions and maintained that "freedom" is applicable only when those belonging to a certain class or classes—e.g., physical restraints imposed upon a person from without—are present.

MacCallum's essay is nevertheless significant because it does emphasize a much overlooked truth: translatability into the *xyz* schema is a necessary condition of the intelligibility of any claim about social freedom. Whenever we use "freedom" in the social sense we ought to specify or at least be able to specify who is free from what conditions to perform what activity.

In a footnote MacCallum mentions the political scientist Felix Oppenheim as one of the few writers who does adopt this rule, but he says that Oppenheim, in his *Dimensions of Freedom*, limits far too sharply the ranges of the *xyz* variables. Since I have just recommended such a limitation myself,

<sup>4</sup> Here I follow the argument of D. P. Dryer in his essay "Freedom," *The Canadian Journal of Economics and Political Science*, vol. 30 (1964), pp. 444-448.



now would be an appropriate time to consider in some detail Oppenheim's work.

#### IV. OPPENHEIM'S DIMENSIONS OF FREEDOM

Oppenheim wants to provide adequate explicative definitions of some of the basic terms employed in political theory, including "influence," "control," "power," and, of course, "social freedom" and "social unfreedom." He wants, that is, to provide definitions which will among other things enhance the clarity and precision of the definienda, thereby reducing the ambiguities and inconsistencies of ordinary usage. Oppenheim's final objective, and what he takes to be the most pressing task confronting political theorists today, is the construction of a conceptual apparatus free of imprecision, obscurity, and contradiction. In what follows I shall argue that his analyses of social freedom and unfreedom do not contribute to this objective, that they have rather obvious counter-intuitive implications and beget serious logical muddles and paradoxes.

The central contention of Oppenheim's chapter on social unfreedom is that there are two ways to render an actor  $X$  ("actor" ranges over individuals and groups) unfree to do  $x$ : one can prevent him from doing it or one can make it punishable for him to do it. I shall first examine the contention that preventing someone from acting constitutes a deprivation of his freedom to act.

Oppenheim begins his analysis by equating " $T$  prevents  $X$  from doing  $x$ " with " $X$  cannot do  $x$  as a result of some action  $y$  of  $T$ ," and then advances the following two definitions:

(a) Where the probability that  $X$  would bring about  $x$  if he chose is extremely small, or where  $X$  theoretically can do  $x$ , but only at the risk of severe deprivation, it is impossible for him to do it, or to say the same thing, he cannot do it.

(b)  $T$  prevents  $X$  from doing  $x$ , if and only if, were  $X$  to attempt  $x$  he would with practical certainty fail to bring it about or suffer extreme deprivation, as a result of some action  $y$  of  $T$  (63).

With regard to (a), Oppenheim acknowledges that he is extending the meaning of impossibility but defends this extension on the grounds it makes the concept more readily applicable to real-life situations (63). But his avowed objective is to provide adequate explicative definitions of terms like "prevention," not to inflate their meanings almost beyond recognition just so they might

enjoy a wider application. That  $T$  prevents  $X$  from doing  $x$  implies that  $X$  did not do  $x$ , but this is not implied by (b): men sometimes manage to defy great odds and accomplish what the polls indicated could not be done, and they also occasionally act despite the risk of severe deprivation.

Using (b) and the belief that preventing a person from acting is one way to render him unfree as premisses, Oppenheim derives the following two theses concerning social unfreedom:

T1. If, upon attempting  $x$ ,  $X$  would with practical certainty fail to bring it about as a result of some action  $y$  of  $T$ , then  $X$  is with respect to  $T$  unfree to do  $x$ .

T2. If, upon attempting  $x$ ,  $X$  would with practical certainty suffer extreme deprivation as the result of some action  $y$  of  $T$ , then  $X$  is with respect to  $T$  unfree to do  $x$ .

Serious problems beset both theses. Looking first at T1, it is rather obvious that the practical certainty of failure to act does not always imply unfreedom to act. Let ' $x$ ' stand for "getting to the church on time" and ' $y$ ' for "stealing  $X$ 's car." And let us suppose that the only way  $X$  can get to church on time is by taking his car. Then  $X$  will almost certainly fail to bring about  $x$  as the result of  $y$ , but we would infer from this not that  $X$  lacks the social freedom to do  $x$  with respect to  $T$ , but that  $T$  has deprived  $X$  of the means (of transportation) necessary to do it. Or let ' $x$ ' stand for "competing in the track meet" and ' $y$ ' for "feeding  $X$  an anaesthetic just prior to the meet." Under these conditions  $X$  will with practical certainty fail to compete in the meet, but we would infer from this not that he lacks the freedom to do so with respect to  $T$ , but that  $T$  has impeded his ability to do so.

Turning to T2, if  $X$  is not aware of the severe deprivation likely in store for him then there seems to be no reason why he should be called unfree in any sense to do  $x$  inasmuch as the prospect of imminent suffering would have no causal bearing on his decision whether to do it or not. What Oppenheim must have meant can be expressed by amending T2 as follows:

T2<sup>1</sup>. If, upon attempting  $x$ ,  $X$  would with practical certainty suffer extreme deprivation as the result of some action  $y$  of  $T$ , and  $X$  knows this, then  $X$  is with respect to  $T$  unfree to do  $x$ .

There are at least two problems with T2<sup>1</sup>. First, it is consistent with the possibility of  $X$ 's doing something he is unfree to do and this is, I believe,

entirely unacceptable. Consider the many individuals who have performed actions, started movements, and initiated reforms knowing full well that their so doing would cause others to inflict severe, sometimes unjustified punishment on them. Should we, do we in fact, say that such individuals were deprived of their liberty to behave as they did? Of course not. When we say that someone has been deprived of his social freedom to do  $x$  at  $t$ , part of what we mean is that he cannot, in the sense of strict impossibility, do  $x$  at  $t$ . Second, even in those cases where  $X$  does not do  $x$  because he knows  $T$  will punish him as a result, it is more accurate and precise to say either that  $T$  has infringed upon his freedom of choice or has undermined his will to do  $x$ . Oppenheim himself emphasizes (in the final chapter) that social freedom and freedom of choice have different meanings and should not be confounded with one another. The former is a political, the latter a metaphysical concept.

A third thesis Oppenheim advances has to do with the relation between social unfreedom and refusing to help a person accomplish some end. Specifically,

T3. If  $T$ 's failure to help  $X$  do  $x$  is at the same time an instance of making it impossible for him to do it, then  $T$  makes  $X$  unfree to do it (70). This is not obviously true, however. Suppose that  $T$  refuses to contribute to  $X$ 's campaign and thereby makes it impossible for  $X$  to win.  $T$  has not rendered  $X$  unfree to win; rather he has deprived him of the financial means necessary to do so. (Of course, without the financial means  $X$ 's liberty in this case was of little meaning, but without the liberty he could not even have run for office.) Or suppose that the government refuses to help the Indians escape from the generally squalid living conditions, and as a result they cannot do so. The Indians should accuse the government of denying them, not the freedom, but the opportunity to improve their lot.

Oppenheim has some more things to say about the relation between prevention and unfreedom when discussing degrees of power and freedom in Chapter 8. His basic thesis reads thus:

T4. If  $T$  prevents  $X$  from doing  $x$ , the degree of  $X$ 's unfreedom to do  $x$  is a function of the utility of doing  $x$  and the disutility of being prevented from doing  $x$  (190). Utility, in turn, is measured by how much or strongly  $X$  wants to do  $x$ . If he is prevented from doing  $x$ , then the greater his

desire to do it the greater the limitation upon his freedom. Oppenheim illustrates this point with the example of a majority voting down a bill which the minority strongly favors: the majority's action infringes upon the minority's freedom to a greater degree than if the issue had been less important to them (190). But here Oppenheim confuses the degree of unfreedom with the degree to which a person or group feels unfree, or feels aggravated and frustrated because he or it is unfree. If  $T$  physically restrains two persons from doing  $x$  he renders both unfree to do it, though if one of them really wanted to do  $x$  while the other did not care we might very well conclude that  $T$ 's action made the former feel more unfree (frustrated, upset, . . .) than the latter. Perhaps feelings can be measured in degrees, but it is extremely doubtful whether unfreedom can be. T4 needs to be modified along the following lines:

T4<sup>1</sup>. If  $T$  prevents  $X$  from doing  $x$ , and thereby renders him unfree to do  $x$ , the degree to which  $X$  feels unfree, or feels aggravated and frustrated because he has been rendered unfree, is a function of the utility of doing  $x$  and of the disutility of being prevented from doing  $x$ .

We have now to consider the second way in which, according to Oppenheim, a person can be rendered unfree to act, namely by making it punishable for him to do so. Oppenheim correctly points out that the expression " $T$  makes it punishable for  $X$  to do  $x$ " can mean one of two things:  $T$  actually penalizes  $X$  for having done  $x$ , or  $T$  has a general disposition to punish  $X$  whenever he does  $x$ . So given Oppenheim's contention that punishability constitutes a sufficient condition of social unfreedom, we can attribute the following two theses to him:

T5. If  $T$  actually penalizes  $X$  for having done  $x$ , then  $X$  was with respect to  $T$  unfree to do  $x$ .

T6. If  $T$  makes it punishable in the dispositional sense for  $X$  to do  $x$ , then  $X$  is to that extent unfree to do  $x$  with respect to  $T$ .

With regard to T5 Oppenheim first clarifies precisely what actual penalization involves. If  $T$  holds that  $X$  did  $x$  and therefore performs some action  $y$  which deprives  $X$ , then  $T$  is said to penalize  $X$  for having done  $x$ .  $T$  deprives  $X$  provided the situation in which  $T$  finds himself as a result of  $y$  is less valuable to him than it would have been without  $T$ 's intervention. Next Oppenheim describes three cases involving the deprivation of

freedom by actual penalization. (1) A congressional investigations committee that finds a witness guilty deprives him of his good reputation and also penalizes him for having expressed certain views. With respect to this committee the witness was, therefore, unfree to propagate these views (75). (2) The majority of a senator's constituency votes him out of office because he supported increased taxation. Therefore he was, with respect to the majority, unfree to support this tax bill (75). (3) A couple's refusal to give their son a lollipop provokes him into a temper tantrum. In effect he is penalizing them for having applied old-fashioned pedagogical principles. Hence with respect to their son this couple was unfree to refuse him the lollipop (76).

But surely neither clarity nor precision (and these are the criteria Oppenheim lays down for the adequacy of explicative definitions) is served by these descriptions. Indeed, they would repel any political theorist intent upon eliminating verbiage, for all we need say is that (1) the witness was unable to propagate his views without incurring punishment from the committee, (2) the Senator could not support increased taxation without losing most of his support, and (3) the parents were not able to refuse their son a lollipop without causing him to become temporarily enraged. Moreover, T5 leads to absurd conclusions. For instance, it implies that a man can be socially unfree to do something he has already done. Of course, if a thief, say, is apprehended and imprisoned he is for the period of his incarceration unfree to steal again, but this is not the same as saying that he was unfree before his capture to break the law. And T5 requires us to hold that if *X* wants badly to be imprisoned then *Y*'s placing him in jail for violating the law does not infringe upon his freedom—in Oppenheim's words "if *Y*'s action does not actually deprive *X*, *Y* does not penalize him in the behavioral sense and hence does not make him unfree in any respect" (73). This is nonsense. Anyone in jail has suffered a loss of freedom, his state of mind notwithstanding.

Turning now to T6, Oppenheim explicates punishability in the dispositional sense thus: *Y* makes it punishable for *X* to do *x* to the extent that, were *X* to do *x*, he would be penalized for having done so as the result of some action *y* of *Y*'s (72). Oppenheim's explication, it should be noted, is not in accord with the ordinary meaning of punishability, according to which *Y*'s making it punishable for *X* to do *x* implies that *X* would be liable to punishment from *Y* were he to do *x*.

At any rate, according to T6 as Oppenheim understands it we should say, for example, that drivers are "officially unfree" to exceed the speed limit to the extent they are actually being fined for doing so (77) and that drivers in the U.S. are officially more free to violate traffic regulations than those in France inasmuch as in France there are fewer patrol cars and it is much easier to argue one's way out of a summons.

But, once again, do we need to say these things? Is it really necessary to speak in such cumbersome, unwieldy, and confusing language? All one has to say, after all, is that, other things equal, drivers are less likely to speed to the extent they are punished for doing so and that because it is considerably easier for drivers in France to speed with impunity than for drivers in America, the former are, on the whole, and other things being equal, less reluctant about doing so. Under most circumstances we have the social freedom to violate traffic regulations; the question is whether we are willing to exercise this freedom in view of, among other things, the likelihood of our being apprehended and punished. So it seems that T6 fails to advance the cause of conceptual perspicuity espoused by Oppenheim and therefore by his own criterion of adequacy for explicative definitions should not be included in any scientific reinterpretation of the vocabulary of political science.

In sum, then, each of Oppenheim's major theses concerning the nature of social unfreedom proves upon close examination to be inadequate. I will now show that his explication of social freedom is also unacceptable.

Like MacCallum, Oppenheim rejects the two concepts view of Berlin. Social freedom is a relation between actors, so any genuine statement about it must be translatable into a statement of the form, "*X* is free to do *x* with respect to *Y*." It is by definition both "freedom from" (being constrained or restrained by somebody) and "freedom to" (do something) (112).

To clarify the nature of liberty Oppenheim introduces the concept of a person's not being free to act. The government makes it neither impossible nor punishable to pay income taxes. Hence we are not unfree to pay them. But the government does make it necessary for us to pay them and therefore we are not, Oppenheim maintains, officially free in this regard. Similarly voters in Belgium, while not unfree to vote, are nonetheless not free to do so either because it is officially mandatory to participate in elections there (111). In

light of this distinction Oppenheim advances the following thesis:

T7. With respect to  $T$ ,  $X$  is not free to do  $x$  to the extent that  $X$  is, with respect to  $T$ , either unfree to do  $x$  or unfree to abstain from doing  $x$  (111). That is, being unfree to do  $x$  entails not being free to do it, and being unfree not to do  $x$  entails not being free to do it. Unpacking T7, we can derive the following four propositions:

(a) If  $T$  prevents  $X$  from doing  $x$ , then  $X$  is to that extent not free to do  $x$ .

(b) If  $T$  makes it punishable for  $X$  to do  $x$ , then  $X$  is to that extent not free to do  $x$ .

(c) If  $T$  makes it mandatory or necessary for  $X$  to do  $x$ —i.e. if he prevents  $X$  from not doing  $x$ —then  $X$  is to that extent not free to do  $x$ .

(d) If  $T$  makes it punishable for  $X$  not to do  $x$ , then  $X$  is to that extent not free to do  $x$ .

Social freedom, according to Oppenheim, is the contradictory, not of unfreedom, but of not being free to act. So  $X$  is free to do  $x$  with respect to  $T$  to the extent that  $X$  is neither unfree to do  $x$  nor unfree to abstain from doing  $x$  (111). Or

T8. If  $T$  does not prevent  $X$  from doing  $x$  and does not make it punishable for him to do it and does not make it necessary for him to do it and does not make it punishable for him not to do it, then  $X$  is to that extent free to do  $x$  with respect to  $T$  (111).

Reflecting on these two theses, it becomes plain (a) and (b) of T7 are both false for the same reasons that prevention and punishability are not sufficient conditions of social unfreedom. As for propositions (c) and (d), they rest on a confusion. While someone who makes it necessary for  $X$  to do  $x$  or punishable for him not to do so may properly be said to deprive him of any freedom of choice,<sup>5</sup> this is not the same thing as depriving him of the liberty to do  $x$ . Indeed,  $X$ 's doing  $x$ , whether involuntarily or not, bespeaks his social freedom to do so. Oppenheim should have said the following:

(c)<sup>1</sup> If  $T$  makes it necessary for  $X$  to do  $x$ , then to that extent  $X$  has no free choice in the matter; and if he does  $x$  against his will, then his behavior is involuntary.

(d)<sup>1</sup> If  $T$  makes it punishable for  $X$  not to do  $x$ ,<sup>6</sup> then  $X$  is to that extent without any free choice in the matter; and if he does  $x$  against his will, then his behavior is involuntary.

Regarding T8, the claim that if  $T$  doesn't prevent  $X$  from doing  $x$  then  $X$  is with respect to  $T$  free to do  $x$  is true; a person's not being prevented from engaging in a certain activity constitutes *by itself* a sufficient condition of his liberty to do so. However, according to T8, to determine whether  $X$  has the liberty to do  $x$  with respect to  $T$  we must know, not only that  $T$  doesn't prevent  $X$  from doing  $x$ , but also that  $T$  doesn't make it necessary for  $X$  to do it, punishable for him to do it, or punishable for him not to do it. If  $T$  does any one of these things, we cannot, in Oppenheim's judgment, call  $X$  free to do  $x$  with respect to  $T$ . But as I have just argued,  $T$ 's making it either necessary for  $X$  to do  $x$  or punishable for him not to do it is relevant only to the question whether  $X$  has any freedom of choice in the matter. And even if  $T$  makes it punishable for  $X$  to do  $x$ , this does not imply that  $X$  is without the liberty to do  $x$ .

We must, then, judge Oppenheim's overall effort to provide empirical, scientifically valid interpretations of social freedom and unfreedom a failure. The next work I shall consider, D. P. Dryer's "Freedom," is devoted in part to evaluating Oppenheim's book, but it also contains some of Dryer's own views on the nature of liberty, which deserve separate consideration.

#### V. D. P. DRYER'S "FREEDOM"

Dryer thinks the meaning of social freedom can best be understood by noticing what would be regarded as diminishing or impairing it. If  $X$  is forced to do  $x$  his freedom is, according to Dryer, impaired. It may also be impaired by  $X$ 's being prevented from doing  $x$ . Now  $X$  cannot properly be said to be forced to do  $x$  unless it is impossible for him to do anything else instead. And, of course, he cannot be said to be prevented from doing  $x$  unless it is impossible for him to do it. Hence, Dryer concludes, Oppenheim was right in emphasizing the relation between unfreedom and impossibility (444).

<sup>5</sup> Strictly speaking,  $T$ 's making it punishable for  $X$  not to do  $x$  deprives  $X$  of the freedom of choice to do  $x$  only if (i)  $X$  knows of the punishment  $T$  has planned for him in case he does not do  $x$ ; (ii)  $X$  knows or believes he will fall victim to this punishment if he fails to do  $x$ ; (iii) this punishment is so severe for  $X$  that he can no longer regard the alternatives to not doing  $x$  as viable.

<sup>6</sup> Here again I would explicate punishability in the manner of n. 5.

Yet, Dryer argues, not every circumstance that makes it impossible for  $X$  to do something is counted as impairing his liberty to act. A snowstorm, even if it prevents a traveler from reaching his destination, cannot properly be regarded as an interference with his freedom, nor can laryngitis be correctly said to render a person unfree to speak.  $X$ 's freedom is diminished only if other men, and not natural events or some circumstance of his own nature, make it impossible for him to do  $x$  (445). In addition, Dryer argues that a person's freedom cannot be impaired unless others make it impossible for him to do something he would choose to do. The liberty of a prisoner who prefers not to leave jail is no more impaired than is that of someone who does not desire to throw himself off a cliff and is protected by a high fence from going over the edge (445). Moreover, a person can suffer a loss of freedom only through the intentional actions or omissions of others (446).

On the basis of the above considerations Dryer advances the following thesis: someone's freedom is impaired through his being forced to do, or prevented from doing something, only if others by their actions or omissions intentionally make it impossible for him to do something he would choose to do (446).

There appear to be at least three problems with Dryer's account. First, it is not at all clear why natural events like snowstorms should be excluded from the class of external impediments capable of infringing upon one's freedom. Is not the assertion "the blizzard rendered  $X$  unfree to continue his travels" perfectly intelligible? Secondly, the belief that a person's freedom cannot be impaired unless others make it impossible for him to do something he would choose to do is quite obviously false. A prisoner lacks the freedom to leave, his state of mind notwithstanding. Dryer confuses the loss of freedom with the feeling of deprivation. His thesis implies that we can make prisoners and slaves free men by conditioning them not to want anything else, and this is a very dangerous idea indeed; from the standpoint of liberty, it is better to be a conscious slave than a happy one.<sup>7</sup> Finally, Dryer's contention that a man's freedom can be abridged only through the intentional actions or omissions of others is at least questionable. It would seem that a man who has been accidentally locked in his room is as unfree to leave as is some-

one who has been intentionally locked inside with him.

Dryer goes on to discuss a second way freedom can be curtailed, through intimidation. Criticizing Oppenheim's account of the relation between unfreedom and punishability in the dispositional sense because it implies  $X$  can be rendered unfree to act by  $Y$  even though he had no knowledge of  $Y$ 's intention to punish him and therefore received no deterrent from such knowledge, Dryer maintains that  $X$ 's freedom to do  $x$  can be diminished by another's punitive response only if he is supplied some motive against doing  $x$  by anticipation of this response (447). Furthermore Dryer contends that  $X$ 's freedom to do  $x$  is not diminished by  $Y$ 's intimidating him unless it is  $Y$ 's intention to visit him with punishment in return for his doing  $x$ . In short, a person's freedom is impaired through intimidation, only if he is supplied some motive to refrain from acting in a given way by the correct expectation it would lead others to do something he doesn't want, and to do it on purpose in return for his acting in that way (447).

It is important to realize that Dryer never questions the idea that liberty can be abridged by intimidation. His major effort is directed toward stating the necessary conditions of this kind of abridgment. But are there not cases of intentional intimidation which are more accurately described in terms of diminished resolve (will, determination) or abated desire than as instances of impaired freedom? Incarcerating or physically enslaving a person is one thing, threatening him is another matter entirely; to label both deprivations of liberty, without further qualification, obscures the important differences between them.

So Dryer's views respecting social unfreedom are on the whole no more acceptable than Oppenheim's. The next work to be examined, C. W. Cassinelli's *Free Activities and Interpersonal Relations*, represents another recent attempt by a noted political scientist to develop adequate explications of important political concepts. I will attempt to show that Cassinelli also fails to illuminate what being unfree means.

## VI. CASSINELLI'S EXPLICATION OF UNFREEDOM

Cassinelli begins by discussing the commonsense and philosophical ideas about two kinds of un-

<sup>7</sup> Here I follow the thinking of B. F. Skinner. See his *Beyond Freedom and Dignity* (New York, 1971), p. 40: "... a system of slavery so well designed that it does not breed revolt is the real threat (to freedom)."

freedom—that of human activities and that of human agents. In his mind almost everyone correctly believes that an activity is unfree when and only when there exist impediments to alternative activities (14), while a person is unfree only when something prevents him from performing an activity (11). Cassinelli gives formal expression to these shared beliefs in the following manner. Since an unfree activity is one the alternatives to which have been impeded, its occurrence can be regarded as highly probable. More precisely,

(1) Activity  $x$  of  $A$  is unfree, if and only if, the probability  $P$  that  $x$  always occurs under the circumstances in which it did occur is high enough to warrant our judgment that the probability of not- $x$  (any activity other than  $x$ ) under these circumstances was negligible (42). A person is unfree, on the other hand, only if something prevents him from performing an activity  $x$ ; and if he is so prevented then his doing something else instead is highly probable—i.e. not- $x$  is an unfree activity. Put formally,

(2)  $A$  is unfree to do  $x$  if not- $x$  (any activity other than  $x$ ) is an unfree activity (42).

There is one final principle Cassinelli defends. It explains, among other things, why we are reluctant to say of a person who has broken the law that he was unfree to do so.

(3) A person cannot be unfree to do something he actually does (34). I have already appealed to (3) on several occasions. It is, in my judgment, self-evident.

But neither (1) nor (2) is self-evident; indeed both are clearly false. The basic difficulty with (1) is that it obliterates the distinction between free and unfree activities, a distinction to which Cassinelli repeatedly refers and which most of us regard as viable and useful (although it cannot be elucidated by reference to the concept of social freedom but is intelligible only when understood to involve the metaphysical issue of free choice). (1) obliterates this distinction because a strong case can be made for the view that all human actions and activities, including even thoughts, desires, intentions, etc., are causally determined by a combination of a person's genetic endowment and past environmental circumstances. Thus given any individual  $A$  the probability  $P$  that he will perform  $x$  under the circumstances in which he has performed it before will always be high enough to warrant our judgment that his doing not- $x$  under

these circumstances is extremely unlikely. (Needless to say it is of crucial importance that the events and facts surrounding and forming the background of a person's behavior be fully and carefully specified, for otherwise, decisive but nonetheless inconspicuous causal factors might be overlooked.) Therefore there is good reason to believe, in light of (1), that all human activities are unfree. Cassinelli has defined free activities out of existence.

Perhaps it is in order to meet just this kind of objection that in the last chapter Cassinelli, contending that the regularity of occurrences definitive of an unfree activity as defined in (1) cannot be interpreted in terms of cause and effect "... because free activities... cannot be interpreted as uncaused activities" (108), introduces the concept of one activity's "resulting" in another. When two activities,  $x$  and  $y$ , are connected by the relationship of resulting, they must be logically independent of one another, occur in temporal sequence, and they must be directly related in the sense that no other event occurs, as it were, between them (104). We should, Cassinelli argues, employ this concept rather than that of causality in analyzing the meaning of "unfree activity." Next Cassinelli identifies the probability  $P$  used in (1) with the degree of credibility of a generalization of the form, "whenever  $x$  occurs,  $y$  occurs," where  $x$  and  $y$  stand in the relationship of resulting, high enough to warrant our dismissal of the statement, " $x$  occurs in the absence of  $y$ ," as having practically no credibility at all (108). Hence according to this altered version of (1) an activity  $y$  of  $A$  is unfree if the generalization, "whenever  $x$  occurs,  $y$  occurs," does not have this degree of credibility—i.e., if we have reservations about the invariability of the connection between  $x$  and  $y$ .

But once again, a strong case can be made in support of the view that, for any activity  $y$  of  $A$ , a highly credible universal proposition of the form, "whenever  $x$  (of  $A$ ) occurs,  $y$  occurs," if not established yet, can and will be discovered as scientific inquiry into human behavior advances. An activity should not be called unfree simply because we do not now know of any law that accounts for it, since there might well be such a law despite our ignorance. Cassinelli's attempt to rescue (1) does not work inasmuch as it provides no principle for determining that some activities can never be subsumed under highly credible generalizations.

Cassinelli replies that this argument again assumes a particular point of view, namely that everything has a cause whether we happen to know it or not, which should only be accepted if it helps us to understand nature. This outlook is less than useful in trying to understand the idea of freedom. Unfortunately Cassinelli fails to explain or show why the deterministic view, which so many scientists have adopted and still utilize precisely because it does help them attain a better understanding of the complex workings of nature, should be repudiated. Of course, determinism is less than useful in comprehending *his* conception of free activities, but he must show that there are reasons for entertaining scepticism with regard to determinism apart from and independent of its implications for his own theory.

There are also a number of difficulties with (2). For one thing, if not- $x$  can be any activity other than  $x$ , the principle implies that no one is free to do anything; for given any activity of  $A$ , it will always be possible to find some other activity, like going to bed every night or speaking to someone during the day, which  $A$  will almost certainly perform. Surely a person's freedom to do  $x$  does not and cannot be made to depend conceptually upon whether his doing something else is highly probable. To render (2) more credible Cassinelli should have specified that "not- $x$ " can designate only  $A$ 's not doing  $x$  (either because he does not try or because he tries unsuccessfully).

But even with this specification Cassinelli's explication of human freedom has counterintuitive implications. For instance, it implies that  $A$  is unfree to lift an elephant or walk on water because his failure to accomplish such feats is highly probable, whereas the correct thing to say would be that  $A$  is physically incapable of performing these kinds of actions.

We must, then, conclude that Cassinelli's analyses of unfree activity and the unfreedom of persons are inadequate and should not be looked upon by other political scientists as "usable elements" for the construction of hypotheses. Whether his explications of power, influence, control, and authority fare any better is, of course, an entirely different matter which cannot be discussed here.

I shall conclude my survey with an evaluation of H. J. McCloskey's essay "A Critique of the Ideals of Liberty" (next section) followed by an examination of the views of S. I. Benn and W. L. Weinstein (Sect. VIII).

## VII. McCLOSKEY'S "A CRITIQUE OF THE IDEALS OF LIBERTY"

So far as I know, of the many articles on liberalism written by H. J. McCloskey over the past 15 years "A Critique of the Ideals of Liberty" is the only one wherein he explicitly makes known his views regarding the nature of liberty. Using Berlin's general distinction between negative and positive liberty as a starting point, McCloskey maintains that there are at least two concepts of negative freedom and four concepts of positive freedom which can be differentiated as follows: (i) non-interference unqualified; (ii) non-interference with the enjoyment of one's rights; (iii) self-determination; (iv) reasonable self-determination; (v) opportunity to enjoy one's rights; (vi) opportunity to seek one's self-perfection. Each of these has, in McCloskey's judgment, a real claim to be called a concept of liberty—i.e., none confuses liberty with some other political ideal—though a couple are more attenuated and less satisfactory than first appearance suggest. While McCloskey also attempts to determine whether any of (1)–(6) represents an ultimate, irreducible ideal, in what follows I shall be concerned solely with his ideas about their genuineness as concepts of liberty.

(i). Negative liberty<sub>1</sub> = Non-interference unqualified. According to this conception to be free consists of being let alone. So a person is free if he is left alone to starve or to die for want of sustenance. Indeed, a person may have perfect liberty in this sense to do many things he desires but be unable to do a single one of them because he lacks the requisite means, ability, power, etc. McCloskey seems to regard this fact as an objection to (i), but as I have argued throughout this paper it is entirely possible that a man socially free to do  $x$  may nevertheless be unable to do it for a variety of different reasons. McCloskey also correctly questions the legitimacy of (i) because not all interferences are normally regarded as interferences with liberty—e.g., one person's hitting another from malice.

(ii). Negative Liberty<sub>2</sub> = Non-interference with Rights. This is historically the most important concept of liberty, according to McCloskey. It is a genuine concept, he argues, because (a) we do charge and explain our charges that our liberty has been violated when our rights are interfered with while (b) we do not complain in the same way when someone prevents us from abusing other



people's property, being cruel to animals, etc. (490). However, there is a problem with (ii) for as McCloskey correctly points out, not every interference with a person's rights is thought to involve a deprivation of liberty—e.g., one man makes another unhappy at insulting him.

It seems to me McCloskey's treatment of this second negative concept is much too sympathetic, that in fact this concept is no more satisfactory or genuine than (i). For one thing, (a) is not always true. Whether we make such a charge depends entirely on the mode of interference employed and its justifiability. If *T* were to prevent *X* from owning property through aspersion, *X* might well charge *T* with slander and take him to court, but it is doubtful whether he would accuse *T* of having invaded his liberty. On the other hand, if *T* were to prevent *X* from owning property by imprisoning him without just cause then *X* would and should complain about an invasion of liberty. And McCloskey is mistaken in assuming the only explanation for (b) is that our freedom has not been violated. Is it not possible we do not *complain* about a loss of freedom when someone stops us from, say, maliciously beating a dog because we realize that under these circumstances the abridgment of our freedom is justified?

Oddly enough McCloskey fails to mention a third negative concept which deserves some consideration, namely, the absence of physical impediments to motion imposed from without. Whether Hobbes's definition needs to be broadened to incorporate mental as well as physical coercion is, of course, open to question; still it should have been included on McCloskey's list of concepts deserving of scrutiny.

(iii). Positive Liberty<sub>1</sub> = Self-determination. McCloskey contends that to be self-determining is to enjoy liberty in an obvious and indisputable sense of that term. In contrast to (i), interferences with alcoholics, drug addicts, and the insane with a view to restoring their autonomy need not, and if successful should not be considered deprivations of freedom in this positive sense. Moreover, concern for freedom as self-determination entails, not simply removing obstacles and compensating for privations, but providing goods like education, fair opportunity, and medical facilities as well (490).

McCloskey identifies self-determination with the general capacity to direct our lives without impediment or hindrance from others (494). But as I argued earlier when discussing Berlin's posi-

tive concept of liberty, having this capacity is neither necessary nor sufficient for a person to be free in the social sense. With regard to McCloskey's contention that any interference which restores a man's self-determination should not be called an abridgment of his freedom, we must remind ourselves of the different forms such interference can take. It might, for instance, involve involuntary hospitalization, in which case there would be an abridgment of liberty but for a good cause. And providing goods like education and fair opportunity is a demand, not of liberty as such, but of justice. These goods are valuable, not because they increase liberty, but because they render its use more valuable.

(iv). Positive Liberty<sub>2</sub> = Reasonable Self-determination. McCloskey argues that this too is a genuine concept of liberty. He writes: "A person is not free if he does not know what he is doing. Hence stupid, unreasonable behavior can, with some point, be said not to be fully free. If a person is not reasonable in his self-direction, if he stupidly yields to impulses, etc., he is not free or fully free" (502).

But as I have maintained all along, if a person actually performs an action, whether he knows what he is doing or not, it is logically most cumbersome to call him unfree in the political sense to do so. Moreover a person's exhibiting reasonable self-determination is not necessary for the truth of the statement that he is socially free to behave in different ways.

(v). Positive Liberty<sub>3</sub> = Opportunity to Enjoy One's Rights. McCloskey admits that people might question whether this is not some distinct political ideal. To show why (v) is a genuine concept of liberty he advances the following two considerations: (a) it has been called liberty in the literature of liberalism and is one of the most important concepts of positive liberty in liberal writings (503); (b) we complain that we are not free if we cannot enjoy our rights, but we do not complain in the same way about coercion which does not impinge upon the enjoyment of our rights (504).

As regards (a), what we want to know is whether (v) should have figured prominently in past liberal writings. We have already seen (Sect. II and Sect. III) that liberty and opportunity ought not be confounded with one another, that they represent two distinct political values. Of course the deprivation of freedom is often accompanied by a loss of opportunity, and vice versa, but to define the one value in terms of the other only



leads to the kind of muddled thinking exemplified in McCloskey's assertion that aids to help the needy and the disabled are conditions for the existence of liberty (504). What we ought to say instead is that such aids are indispensable to the creation of social and economic opportunities for the less fortunate and therefore are necessary for the meaningful exercise of their freedom.

With regard to (b), is it true that when unable to enjoy our rights we always complain about a loss of freedom? This would seem to depend upon the reasons why we cannot enjoy them. A sick man may not be able to enjoy his right to happiness and a poor person his right to life, but neither would likely describe his situation in terms of social freedom. Nor can we infer the genuineness of (v) from the fact that when people are subjected to coercion which does not infringe upon their rights they never complain about a loss of liberty. Perhaps they do not complain simply because in their mind to do so would be either obviously groundless—the use of coercion being entirely justifiable—or inefficacious—no one would listen to their pleas anyway.

(vi). Positive Liberty<sub>4</sub> = Opportunity to be Self-perfecting. Obviously this definition is no more acceptable than (v). The statement that *X* has the liberty to act tells us something quite different from the statement that he has the opportunity to be self-perfecting. McCloskey attempts to defend (vi) by maintaining, among other things, that: (a) the element of free decision and self-determination is an essential component of this account (505); (b) to be self-frustrating or self-destructive is to be unfree in a significant sense (505); (c) if the opportunity to be self-perfecting is treated as a distinct, non-liberty ideal, self-perfection is in danger of being confused with the very different ideal of perfection of self (506).

But a person who has the opportunity to be self-perfecting may not make free decisions and be self-determining. Furthermore, neither self-determination nor free decision is an essential component of political freedom. With regard to (b), a person

who lacks the opportunity to be self-perfecting is not necessarily self-frustrating or self-destructive. And even if self-frustration or self-destruction does result from denying a person this opportunity, it does not follow that he is socially unfree in any respect. Finally, McCloskey never explains or defends (c). Why could not one just as easily argue that law and order ought to be equated with justice since otherwise it is in danger of being confused with racism?

So none of the six concepts considered by McCloskey tells us what liberty is. Although the two negative concepts approximate to a certain degree the meaning of the term, neither expresses a condition necessary for the truth of statements about social freedom. And (iii)–(vi) confuse what it is to be free socially with other goods necessary for the worth of liberty. S. I. Benn and W. L. Weinstein, whose works I shall consider next, at one time agreed with my view that every positive concept of freedom is a counterfeit. But in more recent writings they have moved toward the position of those positive liberals who conceive of freedom as the ability to choose. My principal objective in what follows is to examine the reasons behind their rather drastic and, I believe, wrong-headed conversion.

## VIII. BENN'S AND WEINSTEIN'S WRITINGS

### A. Their Early Views

In the tenth chapter of his 1958 book on *The Principles of Political Thought*<sup>8</sup> S. I. Benn attacks political figures who try to reconcile the spread of state regulation with liberty by defining the latter in terms of opportunity or ability. The main trouble with such positive interpretations of liberty, according to Benn, is that they drain the concept of all descriptive meaning. In his words: "If we want freedom to remain significant as a moral demand, we must be prepared to see it rub shoulders with less exalted ideals, like personal security, full employment, and so on, without wanting it to swallow the lot" (249).<sup>9</sup> Benn recommends that liberty be identified with the absence of restraints

<sup>8</sup> In the Preface the authors, S. I. Benn and R. S. Peters, state that the political ideas advanced in this work are Benn's and the moral ideas Peters'. They also make it clear there is no substantial disagreement between them over any of these ideas.

<sup>9</sup> The emphasis on the importance of conceptual discrimination in political theory, so marked here, also finds expression in Benn's "Some Reflections on Political Theory and Behavioral Science," *Political Studies*, vol. 12 (1964), pp. 237–242. In this later essay Benn writes that the work of clarification which has to be done in the language of politics "... should aim to show how confusions result from the failure to make distinctions . . ." (pp. 241–242). And making the proper distinctions between freedom and other related but distinct values requires that one preserve the subtleties of political discourse. It is regrettable that in his more recent writings Benn comes to defend a conception of freedom which obviously does not preserve these subtleties.

imposed by the power of other men. This definition—call it  $nf_1$ —enables us to distinguish liberty from other related but distinct values, thereby providing the conceptual framework within which we can intelligently assess the impact of social legislation concerning education, health care, taxation, etc., on human well-being.

While Benn's objective is certainly commendable, his proposed definition of liberty is inadequate. An expensive education, job discrimination, physical disability, and personal qualities like shyness or laziness may on occasion serve as restraints to human behavior, but they are not usually regarded as interferences with social freedom. I have argued all along that only certain kinds of restraints or obstacles are properly thought to render a man socially unfree to act. The important point to note here, however, is that there is nothing radically defective about  $nf_1$ , that it can be amended to meet this objection. One should not infer that, because  $nf_1$  in its present, unqualified form is unacceptable, negative liberalism must be abandoned.

In his 1965 essay, "The Concept of Liberty in Nineteenth Century English Political Thought," W. L. Weinstein contends that our everyday language is sufficiently varied and precise to eliminate the need to change the basic, negative meaning of "being free" or to equate it with other things (160). What is its basic meaning? According to Weinstein liberty is the absence of interference or coercion by other human beings (146). Contrasting this definition—call it  $nf_2$ —with other theories of freedom which have been advanced by positive liberals, Weinstein makes the following points. (a) It is a mistake to equate liberty with any kind of power or capacity to act; when we ask whether a man is free to do something, our concern is not with whether he has the power to do it. (b) Nor does the concept of self-development throw any light on what "liberty" means; rather it constitutes one possible goal or purpose of being free (151). (c) Social benefits like low cost housing and welfare are means to self-development and therefore will undoubtedly render man's freedom more valuable, but they are not constitutive of his freedom (153). (d) Finally, liberty should not be confused with self-knowledge. A prisoner, after all, might come to know himself very well in jail, but this does not entail that he thereby becomes more free (159). Self-knowledge is valuable, not because it increases a man's freedom, but because it enables him to make wiser,

more rational choices. Weinstein is surely right in drawing these distinctions.

At the close of his essay Weinstein briefly discusses a couple of cases where it is, in his mind, difficult to determine whether  $nf_2$  is applicable. If an oligarchy sets out deliberately to keep the rest of society illiterate and ignorant by not providing schools, the similarity between this policy and direct legal prohibition goes beyond their achieving the same result. So a good case can be made for accusing the oligarchy of violating men's freedom to attend school. And if  $X$  is deliberately weakened or "conditioned" (as in brainwashing) by  $T$ , the resemblance to coercion is so close that we may be justified in calling  $X$  less free and not merely less capable of using his freedom (162).

But while there may be doubt whether these cases involve deprivations of freedom as Weinstein defines it, it is clear they do not entail threats to freedom properly conceived. The oligarchy is attempting to deny the rest of society, not the freedom, but the opportunity to receive an education;  $T$  deprives  $X$  only of the ability to think for himself. Weinstein's definition, like  $nf_1$ , needs to be amended in such a way that only certain kinds of interferences and coercion are classified as dangers to freedom. Not all interferences curtail freedom (see the discussion of McCloskey's essay in the last section). Nor does the use of coercion always result in a loss of freedom: for one thing persons sometimes pursue activities in spite of warnings or threats not to do so; and even when coercion does work, the expressions "deprivation of will (resolve, determination)" and "deprivation of desire (inclination)" often provide a more accurate description of what it effects.

#### *B. Benn's and Weinstein's Turnabout*

Mass persuasive techniques like commercial advertising and political propaganda, which aim to manipulate opinion and shape preferences, provide the subject matter of Benn's 1967 article on "Freedom and Persuasion." While admitting that these influences are less successful than many early alarmists of the 1950s suggested, Benn contends they nevertheless do constitute a definite threat to freedom, their subtle, often inconspicuous techniques notwithstanding. But freedom conceived as the absence of interference or restraints by other men is insufficient to account for this threat and for the uneasiness contemporary liberals experience when confronted with it. Advertisers and propagandists do not, after all, restrain or interfere with

the consumer; on the contrary they seek to create a willing accord between the buying public and themselves. Consequently a re-examination of concepts like liberty and interference is in order. Benn's objective is "... to extend the application of the classical concepts of liberalism so that they can function coherently in the discussion of persuasive 'techniques'" (260).

The concept of freedom the liberal needs to form his objections to certain of these persuasive techniques is none other than rational self-mastery (262). For in many cases commercial advertising and political propaganda make it extremely difficult for the exposed person to make up his own mind concerning different products, issues, and life-styles. Of course not all persuasive techniques jeopardize freedom in this sense. Persuasion by arguments which invite and respond to criticism, for example, is perfectly consistent with a person's determining his own mode of existence in his own way. But there are non-rational persuasive techniques which are not consistent with this value, and Benn devotes much of "Freedom and Persuasion" suggesting criteria for identifying them.

The basic difficulty underlying Benn's conversion pertains to his belief that the classical concepts of liberalism *must* be re-examined and extended with a view to making them applicable to opinion-manipulation. Why is it necessary for the liberal or anyone else bothered by mass persuasive techniques to express his criticisms in terms of liberty; especially when there are other concepts available which more accurately describe the threat posed by these techniques? Surely all we need say is that the widespread use of commercial advertising and political propaganda threatens man's ability to think for himself and consequently makes it very difficult for him to realize the full worth of his liberty. The persuasion industry imperils human rationality, not social freedom. Stating the problem in this manner reflects a commitment to conceptual clarity; it preserves the subtleties of political discourse without betraying any liberal principles.

In "Being Free to Act, and Being a Free Man" published in 1971, Benn and Weinstein reject  $nf_1$  and  $nf_2$  for the following reason. Consider the cases of two men:  $X$  cannot purchase eggs because he hasn't enough money;  $Y$  does the only thing permitted by law, but it is something he would have

done anyway. According to these negative definitions  $X$ , having been impeded or restrained from purchasing eggs, must be called unfree to do so, while  $Y$ , who is not impeded or restrained, must be considered free to disobey the law. But it is odd to call  $X$  unfree to buy eggs, inasmuch as the option of having free eggs was never open to him in the first place (202). And it is also odd to call  $Y$  free to disobey the law, since this option is not really available to him (197).

These cases suggest that deciding between available alternatives, not the absence of impediments or restraints, lies at the heart of freedom. Then Benn and Weinstein contend that to be free means "... to be able to choose among available courses bearing in mind their expected consequences both good and ill" (198). Freedom should be identified with the non-restriction of available options (201). This definition enables us to avoid calling  $X$  unfree to purchase eggs; charging a price for eggs, though it impedes or restrains  $X$  from buying them, does not close any option that once was available. And it enables us to call  $Y$  unfree since the law in question does limit otherwise available courses of action—whether  $Y$  would have chosen them or not makes no difference.

The first and most obvious problem with Benn's and Weinstein's position is that the two formulations they give of their proposed definition of freedom are not equivalent. It is one thing to say that  $X$  has the general ability to choose from available options, quite another to say that there are no restrictions on these options. The first statement refers to a specific property belonging to  $X$ , the second does not. Furthermore neither formulation adequately explains what liberty is. To be *free* in the political sense does not entail having the capacity to choose, nor does this capacity entail political freedom. And there are a number of ways to restrict a man's available options without invading his liberty—e.g., by knocking him unconscious or depriving him of the necessary means to act.<sup>10</sup>

Benn and Weinstein clearly would have been better off had they attempted to amend the standard negative definition of liberty in such a way that the lack of financial resources, for example, would be excluded from the class of impediments or restraints substitutable for the  $y$  term in the triadic schema, " $X$  is free from  $y$  to do  $z$ ." As I

<sup>10</sup> For a more detailed discussion of this point see my "Freedom as the Non-Restriction of Options," *Mind*, vol. 83 (1974).

pointed out earlier, there are difficulties with  $nf_1$  and  $nf_2$ , but they do not portend any radical defect with either definition.

A satisfactory repair job would result in a definition much like the following: liberty is the absence either of physical restraints imposed (by man or natural forces) on or around a person which prevent him from moving except within a limited space, of moral coercion employed so effectively and/or unexpectedly that it serves as a sufficient causal condition of the victim's not acting in a specified way,<sup>11</sup> or of institutional arrangements which make it impossible to undertake certain kinds of activities—e.g., in a country which prohibits divorce for any reason, no one is socially free to obtain a divorce.

Applying my amended negative definition to Benn's and Weinstein's examples,  $X$  should be called free to purchase eggs even though he is without the financial means to do so, while  $Y$  should be described as free but lacking the desire to disobey the law. Philosophers and political scientists concerned with the construction of a conceptual apparatus free of imprecision, obscurity, and contradiction ought to take a careful look at this revised interpretation of negative liberty.

#### IX. CONCLUDING REMARKS

The definition I have just offered is meant only to clarify an understanding of liberty. It cannot serve as a norm for adjudicating conflicting claims to liberty. Now there are those who would regard this as a real deficiency of my account. But it seems to me that questions concerning the conditions under which social freedom should and should not be interfered with belong to substantive political philosophy and can only be answered by a theory of justice such as has recently been proposed by John Rawls.

Indeed, in my judgment many of the positive conceptions of liberty which philosophers have defended in the past should be construed, not as definitions, but as principles of justice constituting

criteria for assessing claims to liberty. Consider, for example, the conception of liberty as rational self-mastery. Only confusion results from identifying liberty with this value. But it is possible, even credible, to equate rational self-mastery with a goal or end of liberty. Thus we have the principle of justice, that everyone has a prima facie right to achieve rational direction over his life, and this principle can be used, along with others, in resolving disputes concerning the legitimate distribution of liberty. Conceptions of liberty in terms of opportunity should be treated similarly. And, of course, we need to insure that persons are not subject to arbitrary or purposeless restraints, but again the best way to do this consists, not of defining liberty in such a way that only rational restraints can be said to abridge it, but of embodying in principles of justice prohibitions against unfair or pointless infringements upon liberty.

Rawls's *A Theory of Justice* contains the most thorough and insightful contemporary inquiry into the relation between liberty and justice, so it would be appropriate for me to conclude by commenting briefly on the concept of liberty Rawls employs in this work. He correctly observes that: (a) the general description of liberty must always specify who is free (or not free) from this or that constraint (or set of constraints) to do (or not to do) so and so (202); (b) things like poverty, ignorance, and lack of means diminish the worth of liberty, not liberty itself (204); (c) disagreements relating to the values of different liberties can only be settled by a theory of justice (202). But Rawls mistakenly believes that legal duties and prohibitions as well as coercive influences arising from public opinion and social pressure deserve to be counted among the constraints which do limit one's liberty (202). As I have argued, only certain kinds of laws are properly deemed infringements upon social freedom, while public opinion and social pressure can undermine a person's will or resolve, but not his liberty, to act. Rawls's conception of liberty needs to be qualified in much the same way Benn's and Weinstein's early definitions did.

*The University of Santa Clara*

*Received June 14, 1973*

<sup>11</sup> Moral coercion must be so qualified to discriminate interferences with freedom from what are more aptly described as deprivations of desire, will, and purpose.

## BIBLIOGRAPHY

- BENN, S. I., "Freedom and Persuasion," in *Australasian Journal of Philosophy*, vol. 45 (1967), pp. 259-275.
- , "Some Reflections on Political Theory and Behavioral Science," in *Political Studies*, vol. 12 (1964), pp. 237-242.
- , and PETERS, R. S., *The Principles of Political Thought* (New York, 1958), ch. 9 and 10.
- , and WEINSTEIN, W. L., "Being Free to Act, and Being a Free Man," in *Mind*, vol. 80 (1971), pp. 194-211.
- BERLIN, Isaiah, *Four Essays on Liberty* (Oxford, 1969).
- , *Two Concepts of Liberty* (Oxford, 1958).
- CASSINELLI, C. W., *Free Activities and Interpersonal Relations* (The Hague, 1966).
- COHEN, Marshall, "Berlin and the Liberal Tradition," in *Philosophical Quarterly*, vol. 10 (1960), pp. 216-228.
- DILMAN, Illam, "The Freedom of Man," in *Proceedings of the Aristotelian Society*, vol. 62 (1961-62), pp. 39-62.
- DRYER, D. P., "Freedom," in *The Canadian Journal of Economics and Political Science*, vol. 30 (1964), pp. 444-448.
- GILL, John, "The Definition of Freedom," in *Ethics*, vol. 82 (1971), pp. 1-20.
- KAUFMAN, A. S., "Professor Berlin on 'Negative Freedom'," in *Mind*, vol. 71 (1962), pp. 241-243.
- MACCALLUM, Gerald, "Negative and Positive Freedom," in *The Philosophical Review*, vol. 76 (1967), pp. 312-334.
- MCCLOSKEY, H. J., "A Critique of the Ideals of Liberty," in *Mind*, vol. 74 (1965), pp. 483-508.
- OPPENHEIM, Felix, *Dimensions of Freedom* (New York, 1961).
- RAWLS, John, *A Theory of Justice* (Cambridge, 1971).
- SPITZ, David, "The Nature and Limits of Freedom," in *Dissent*, vol. 8 (1961), pp. 78-83.
- WEINSTEIN, W. L., "The Concept of Liberty in Nineteenth Century English Political Thought," in *Political Studies*, vol. 13 (1965), pp. 145-162.
- WHITE, D. M., "Negative Liberty," in *Ethics*, vol. 81 (1970), pp. 185-204.

## II. REASONS, WANTS, AND CAUSES

DON LOCKE

ONCE upon a time, not so very long ago, it was a commonplace of the theory of action that there is a crucial distinction to be drawn between reasons and causes. For clearly there is an important difference between those things which an agent does because he has some reason to, and those things which he does—or we might more accurately say, which happen to him—as the result of some cause which affects him. And, it was suggested, this difference, one which seems central in any discussion of the freedom of the will, can best be clarified and explained by concentrating on the distinction between a reason and a cause.

Nowadays, however, it is every bit as much a commonplace of the theory of action that reasons are themselves but one kind of cause. In many contexts “reason” and “cause” seem interchangeable; the smoldering coal which falls from the grate onto the hearthrug may be both the reason why the house catches fire, and also the cause of that event. And where a man acts for a reason, he acts as he does because he has that reason; if he did not have that reason he would not act as he does. It is, at the very least, obscure why we should insist that that reason is nevertheless not the cause of his action. Of course reasons do differ from other causes, like physical forces or the fear of death which causes a man to tremble, but these other causes also differ considerably among themselves.

At this point the question of whether reasons are causes is in danger of becoming a purely verbal debate, a matter of what restrictions we are or are not prepared to build into the word “cause.” The substantive issue will be just how reasons do differ from the standard varieties of cause, and especially whether these differences are such as to show that actions done for reasons differ in nature or status from events brought about by (other) causes. To answer that question we need first to investigate the nature of reasons themselves.

### I

There is a familiar distinction between *a* reason for an agent’s acting in a certain way, *the* reason

why he acts in that way, and *his* reason for so acting. I am here concerned only with reasons of this third sort, reasons which are the agent’s reasons, which I shall call “agent-reasons” when they need to be distinguished from reasons of other sorts. Such agent-reasons will include not only the agent’s reasons for doing what he does, but also reasons which he has but does not act on, whether because he does not act in that way at all, or because he acts in that way for some other reason; but they do not include reasons which others see the agent as having but he does not.

Agent-reasons are, most commonly, beliefs which the agent has, things which he believes to be so, under which I include not only matters of fact but also, if these are different, matters of value. My reason for doing something may be that, as I believe, this man has insulted me; or that, as I believe, it is going to rain; or that, as I believe, I ought to do it. Such considerations provide reasons for an agent only insofar as he believes them to be so: if I do not believe that this man has insulted me then it can hardly be that I slap his face because, for the reason that, he has insulted me—even if the fact is that he has insulted me. Indeed the belief need not even be true to provide the agent with a reason: it may be that I am mistaken in thinking that he has insulted me, but so long as I think, i.e., believe, it, then that is my reason, good or bad as the case may be, for slapping his face. Notice here that it is the fact *that* I believe it which explains my action, and hence provides the reason why I act; but my reason for acting, the agent-reason, is not that I believe it but *what* I believe, the content of my belief, viz., that he has insulted me. We shall be returning to this important point of difference between *the* reason and *my* reason.

A second group of agent-reasons consists of such things as the agent’s hopes, fears, doubts, wishes, feelings of all sorts, desires, needs, etc. Thus my reason for doing something may be that I am uncertain whether I have turned the gas off; or that I need someone to talk to; or that I feel bored, or feel cold, or feel something bite me. These constitute a distinct group only in that there seems

no question of the agent's failing to realize that he hopes, fears, feels, etc., and hence it seems inappropriate to speak of the agent as believing that he has these hopes, fears, feelings, etc. But if it were possible to feel uncertain or feel bored or feel a bite without realizing that you did, then this second group would reduce to the first. They would be reasons only insofar as the agent believed that he felt uncertain, felt bored, or felt a bite; his reason would be that, as he believes, he feels uncertain, or that, as he believes, he feels bored, or that, as he believes, he feels a bite.

Let us therefore concentrate on the first group, where the agent's beliefs provide him with his reasons for acting. Now if some consideration is to be, for me, a reason to do something, more is necessary than that I believe it: I believe that grass is green and that the sky is blue, but these are not, for me, reasons for doing anything, not at the moment at any rate. So what makes a particular belief a reason for that agent to do something? A first answer might be that a belief constitutes an agent-reason just insofar as it is a belief which tends to make the agent act in some way. Events are causes just insofar as they produce effects and similarly, perhaps, beliefs are reasons just insofar as they lead to action, or would lead to action other things being equal.

But this cannot be quite right, since there are beliefs which tend to make us act in various ways, which are nevertheless not *our* reasons even though they are *the* reasons why we act in those ways. For example, it may be that if I believe that a student has a low opinion of my teaching I tend to give him low marks in his examinations. I am assuming here that I am an incompetent examiner, not a wicked one; that is, although I believe that the student has this low opinion of my teaching, I am also convinced that this has nothing to do with the marks I give him. In such a case my belief that he has this low opinion of my teaching is *the* reason why I mark him down, or at least one of the reasons, but it is not *my* reason.

What, then, would make it my reason for marking him down, thus turning me from an incompetent examiner into a wicked one? It seems that I must not only believe that he has this low opinion of my teaching, but also realize that this is why I am giving him low marks. But is simply recognizing that something is *the* reason why I act sufficient to make it my reason? If I realize that the reason why I speak rudely to this man is that he reminds me of the school bully who terrorized

my youth, does that mean that this is my reason for speaking harshly to him? I might rather admit that I have no reason for speaking so harshly to this particular man, that my behavior is to this extent irrational, though I can understand why I do it.

No, what seems to be necessary for me to have a reason for doing something, for it to be my reason for doing it, is that I accept it as such. If my reason for marking the examinee down is that he has a low opinion of my teaching, then it is not merely that I believe that he has this low opinion, but also that I believe that this is a reason for marking him down—and it is precisely this latter belief which shows me to be a wicked, not merely an incompetent, examiner. More generally, for my belief that *p* to be, for me, a reason for doing *x* I must believe that *p* is a reason for doing *x*. This may appear circular but in fact is not, for this latter belief is a belief not that something is *my* reason, but that it is *a* reason. And even this apparent circularity can be avoided by offering some analysis of what it is to believe that *p* provides a reason for doing *x*, e.g., that it is to believe that if *p* is the case then *x* is the thing to do, if you can, and have no other, better reason not to do *x*.

To have a reason, then, is to believe something. More accurately, where the reason is itself a belief, something the agent believes, then it is to believe two things: if *p* is, for me, a reason to do *x*, then I believe that *p* is the case; and I also believe that (the fact of) *p* provides a reason for doing *x*. This second belief must also be present in the case of reasons which are not beliefs: hopes, fears, doubts, wishes, feelings of various sorts, desires, needs, these will provide an agent with reason to act in various ways only insofar as he believes that they do. The fact that I feel cold is, for me, a reason to light the fire only if I believe that it is; feeling cold can hardly be my reason for lighting the fire, or a reason for me to light the fire, if I do not believe that it is a reason for doing so. Thus the crucial feature of agent-reasons, of whatever sort, is that the agent regard them as reasons.

This feature of agent-reasons, that to qualify as an agent-reason a belief must be accepted by the agent as a reason, is important again when we consider what it is to act *for* a reason. To do *x* for a reason is not simply to act because you have some belief which constitutes an agent-reason; it is to act because you *accept* that belief as a reason

for doing it. An example may make this clear. I am sitting on a rock on the hillside, eating my sandwiches, when suddenly I hear a hissing sound beneath me. Thinking it is a snake I leap to my feet, scattering my lunch in the dirt. Thus I jump up because I believe there is a snake coming at me. This belief is (part of) *the* reason why I jump, but is it *my* reason for jumping? The fact that my jumping is an automatic, almost instinctive response to the hissing sound suggests that I don't jump for a reason at all; I just jump. Yet it may well be the case that I regard the possible presence of a snake as a perfectly good reason for getting out of the way. That is, I accept it as a reason for jumping up; I do have a reason for jumping up as I did. But insofar as I would have jumped up like that whether I had had that reason or not, that reason is not my reason for jumping. Of course I would not have jumped if I had not believed there was a snake coming at me. But it may be that I would have jumped even if I did not regard that as a reason to jump up in this hasty and thoughtless way. Or it may be that as well as having a reason for jumping up, I also had a reason for not jumping up, viz., that it meant losing my lunch, and that so far as reasons go I regard the second as outweighing the first—that is, I care more about my lunch than I do about the snake—but even so my natural and automatic response is to jump, and spill my sandwiches. In these cases I do have a reason for jumping, but I do not jump *because* I have that reason, and hence it is not my reason for jumping, I do not jump *for* that reason.

It is instructive to note how this analysis deals with the puzzling case of rationalization, where a man quite sincerely offers something as his reason, when we know it is not the reason why he acts. Clear examples are provided by cases of post-hypnotic suggestion: under hypnosis the subject is instructed to take off his shirt at a given cue; he is brought out of hypnosis and the cue is given; the subject complains of the heat, opens a window, and takes off his shirt. Here the agent gives as his reason for taking off his shirt the fact that the room is hot, but we see this as simply rationalization. It may be his reason, but it is not the (real, true) reason. Indeed there is some temptation to say that it is not even his reason, that strictly he had no reason at all, he simply invented one; and certainly, even if it was his reason, he did not do it for that reason. I believe that the analysis I have offered explains why there is a case for denying

that it is his reason, why there is a case for insisting that it is his reason, and why, even if it is his reason, it is not the case that he does it for that reason.

According to our analysis for it to be an agent-reason, a reason of his, the agent must believe both that the room is hot, and that this provides him with a reason to take his shirt off. So the more we are inclined to say that, despite what he says, he can't believe that the room is hot—it is quite cold, say, and he must know that it is—the more we will be entitled to deny that that is even his reason for taking his shirt off. Or the more we are inclined to say that, despite what he says, he can't really regard that as a reason for doing what he does—he just isn't the sort of person who would regard the heat of the room as a reason for taking his shirt off in the middle of a polite dinner party—the more we shall be entitled to deny that this is his reason for taking his shirt off. But if on the other hand there is no reason to deny that he believes that the room is hot, and also no reason to deny that he regards this as a reason for taking his shirt off, there will be no reason to deny that the room's being hot is his reason why he took his shirt off, even though it is not the reason. Nevertheless, it will still not be correct to say that he did it for that reason, simply because although he believed those things it was not because he believed them that he acted as he did. If he did take his shirt off because he had those beliefs, then this is not a case of rationalization after all, though it may be a case of overdetermination.

However, our analysis does not deal so easily with another puzzling but equally possible case, where an agent acts irrationally even by his own lights, in that he does something for reasons which he himself considers inadequate. For example, a man might give up his job, for some fairly trivial reason, although agreeing that he has more or better reasons for keeping it. There is no difficulty here if he gives it up arbitrarily, for no reason at all. But our account does not allow that he might have a reason, act for that reason, and yet not consider it as a reason, or at any rate a good enough reason, for acting as he does. One way of handling this case would be to say that, strictly, he does not act for that reason which he agrees to be inadequate. Rather, he simply wants to give up his job, though he admits he has no very good reason to do so, and it is his wanting to give up his job, not the various inadequate reasons, which explains his doing what he does. Now so far I have tried to



explain both what it is to have an agent-reason, and what it is to act on an agent-reason, without making any reference to wants. This is entirely deliberate. The notion of wanting is a notoriously difficult one, with wants quickly becoming some elusive and unanalyzable datum of inner experience. But we are here faced with an example where a man's reason for acting seems not to be some consideration which he regards as a reason, as our analysis lays down, but simply a want. We can put it off no longer: it is time to face up to wants.

## II

What, then, is it to want to do something? One of the difficulties here is that the one word "want" can, if necessary, range over a vast variety of phenomena, from over-powering passions to simple intentions. But in the wide and generous sense which "want" tends to have in the theory of action, and which I shall be concerned with here, it has nothing specifically to do with passions or desires; a man can be said to want to do something simply insofar as he does it, or would do it, voluntarily and intentionally. In this formal sense, as I shall call it, to want to do something is not, or not necessarily, to be in the grip of some grand passion; it is not even to favor or approve or look forward with pleasure to doing whatever it is. It may be that I want to get up at five in the morning, because I have a train to catch, without bursting with desire to do it, or even favoring or approving or looking forward to getting up so early. It is simply that that is what I am going to do, if I can, and if no better reason for staying in bed emerges.

This suggests that wanting, in our formal sense, can be reduced to having reasons: to want to do something, in this sense, is simply to have some reason for doing it; the more reason or the more reasons I have for doing something, the more I want to do it; and what I most want to do will be that which I have most reason or reasons to do. But this reduction of wanting to having reasons is open to challenge from two directions. On the one hand it may be objected that obviously a man may have very good reason to do something, e.g., go to the dentist, and yet not want to do it. But this, like my getting up at five to catch a train, is a case where I both have some reason for performing a certain action, and also have some reason for not performing it; and hence, according to

our analysis, a case of both wanting to do it and wanting not to do it. I want to go to the dentist because my tooth is hurting, but I also do not want to go to the dentist, because I know he will make it worse before he makes it better. We might say of such a case that what I *really* want is to get my toothache cured, and I go to the dentist, even though this is not something I really want to do, because it is a way of getting what I really want. This seems to mean that the fact that it is going to the dentist is not, *in itself*, a reason for doing what I do, whereas the fact that it is, hopefully, a way of getting my toothache cured is, for me, a reason for doing it.

On the other hand it may be objected that a man may want to do something, and yet have no reason for doing it. The sun is shining; I feel like singing; I burst into song. Why do I sing? I have no reason, I just feel like it, I simply want to sing. But what this means, I think, is not that I have strictly no reason at all, but that I have no further reason—or as we sometimes say, no particular reason—for singing, no reason beyond the fact that doing this is singing. I sing not because it will draw attention to myself, or reveal my melodic voice, or show you what a jolly fellow I am, but just because this is singing, after my fashion. So although this is a case where I have no further reason for what I do, it is not strictly a case where I have no reason whatsoever, and hence not a case where the agent wants to do something without having any reason to do it.

Of course it has to be admitted that it is highly artificial to say, in such a case, that my reason for singing is that what I am doing is, as I believe, singing. The natural idiom is to say that I have no reason. But whether we say that I have no reason, or that I have a reason of this peculiarly degenerate sort—and does it matter so much what we happen to *say*?—I think it can be seen that this case is continuous with the others, where we unhesitatingly speak of the agent as having reasons for what he does. Even in this case my wanting to do it consists in my satisfying the conditions for having some reason to do it.

We are now in a position to deal with the case of the man who acts irrationally, doing something which he recognizes he has no adequate reason to do. What happens here, we suggested, is that the agent does it simply because he wants to, in spite of having no adequate reason to. But according to our present analysis of wanting, to say that he wants to do it is to say that he has some reason to.

The apparent contradiction is resolved, once again, by pointing out that to say he has no reason to do it is to say that he has no further reason to do it, which does not conflict with the fact that he wants to do it in itself or for its own sake, where his reason for doing it is simply that it is the action that it is. Thus the man who gives up his job, irrationally, though he realizes he has no good reason to, does act for a reason after all. His reason for doing it is simply that it is giving up his job; and, as with all reasons, he acts on that reason only insofar as he believes that the fact that it is giving up his job is a reason for doing it. This is compatible with his having no other belief which, he believes, provides him with a reason to give up his job. That is, it is compatible with his having no other, no further reason for doing what he does.

Now there are two familiar theories about the connection between reasons and wants. One is that reasons presuppose wants, in that part of what it is to have a reason is to want something. Thus Donald Davidson,<sup>1</sup> for example, has suggested that where an agent does something for a reason, he will both believe that he is performing an action of a certain sort, and want to perform an action of that sort. The other well-known theory is that reasons require wants if they are to be effective in action. Thus Hume, for example, has argued that "reason is and ought only to be the slave of the passions," that "reason alone can never be a motive to any action of the will." This is to say, apparently, that a man will not do what he has reason to do unless he also wants to do it. It is perhaps worth pointing out that although these two theories are not exactly inconsistent, the second fits oddly with the first, in that it would be otiose to insist that reasons require wants if it is also the case that reasons presuppose wants. We can regard the Humean theory as weaker than the Davidsonian one, in that Hume's claim might well be true and Davidson's false, though not vice versa.

So far, at any rate, it seems to be the stronger, Davidsonian, claim that is true, though trivially so, inasmuch as to have a reason is to want, and vice versa. It isn't that you need a want to turn a belief into a reason; rather a belief's being a reason is precisely what a want is. This not only explains the connection between reasons and wants without endorsing the puzzling split between

reason and the passions which Hume urged upon us; it also avoids any duplication or dualism in the explanation of action. All we need, in fact, to explain people's actions are their beliefs: people act as they do because they believe what they do.

But unfortunately our whole discussion so far rather over-estimates man's rationality. We have ignored the possibility of *akrasia* or weakness of will which despite its usual location in discussions of moral philosophy is essentially a problem in the theory of action, as Davidson has pointed out (albeit in an anthology of essays on moral philosophy).<sup>2</sup> The fact of it is that a man may have very good reason to do something, recognize that he has, and yet not do it. No doubt there will be some reason why he does not do what he believes he has very good reason to do, but this reason need not be his reason, and so may not be covered by our qualifying clause that he will do  $x$  "unless he has some other, better reason not to do it." For example, I have had a bad day and come home feeling fed up and mildly depressed, certainly listless and lethargic. I am invited out to meet Mr. Jones whom I have always expressed the greatest interest in meeting, whose acquaintance I am sure I shall both enjoy and profit from. I have no very good reason for staying home, certainly not compared with all the good reasons I have for meeting Mr. Jones, but I can't be bothered, so I stay at home. My depressed and listless state of mind may be the reason why I stay at home, but it is not my reason for staying home. Or even if it is, for me, a reason for staying home, at the same time I realize I have more reason to go out. Nor am I simply being short-sighted, letting short-term considerations outweigh longer-term ones. I recognize that from any point of view I have much more reason to go out and meet Mr. Jones. But I do, deliberately, voluntarily, and intentionally, what I realize I have more reason not to do.

There are those who deny that *akrasia* can occur; they will insist that the case given has been misdescribed in some way. For them there will be no problem, and everything that has been said so far can stand. But unhappily *akrasia* seems to be all too possible, and our account of wanting has to be altered to take account of it. For it cannot be handled simply as a case where the agent has reasons both for and against doing something, and

<sup>1</sup> "Actions, Reasons, and Causes," *The Journal of Philosophy*, vol. 60 (1963), p. 686.

<sup>2</sup> "How is Weakness of Will Possible?", in Joel Feinberg (ed.), *Moral Concepts* (London, 1969), p. 102n.

hence both wants to do it and wants not to do it. The puzzling feature is that it is a case where the agent has, and admits he has, more reason to do *A* than *B*, but yet wants to do *B* rather than *A*. This is a possibility which our account has failed to leave open. Indeed it seems that wanting is precisely what has been left out of the picture so far. A man may have every reason to do something, but not want to do it; and if he does not want to do it, then he won't do it. Perhaps it was Hume after all who told the truth about the connection between reasons and wants: if we are to explain human action we need not only belief but also something quite different, desire.

There now seems to be at least this difference between having a reason and wanting, that a want, to be a want at all, must at least influence action, whereas a reason need not. This is not to say that all wants result in action, for that is plainly false, but it is to say that if an agent wants to do something then he will at least be inclined to do it, in that he would do it other things being equal, i.e., given no obstacles or countervailing factors, and in particular given no countervailing wants. So given that wants, qua wants, incline us to behave in certain ways, might we not say that to want to do *x* is to have some "effective" reason for doing *x*, i.e., some reason which actually inclines the agent to do *x*? More accurately, perhaps, we might say that a want is a reason which is effective as a reason, because I accept it as such. In the case where I hear a hissing sound and jump up because I think it is a snake, I have a reason for what I do, and that reason is certainly effective. But is it necessarily the case that I wanted to jump up like that? Perhaps I would have jumped whether I had wanted to or not. The point here seems to be that I want to jump only insofar as I do so for some reason which I accept as a reason, or at any rate as a better reason than my reasons for not jumping. If I jump even though I do not regard the possible presence of a snake as a good enough reason for jumping and losing my lunch, then, it can be agreed, I do not want to jump; I jump in spite of myself.

Thus the case of *akrasia* forces us to modify our account of wanting, but while more complex, the modified account has much the same advantages. We do not need to introduce wants in contradistinction to reasons. To want to do something is

to have some reason for doing it, some reason which actually affects our action, in that it leads to action or would lead to it, other things equal, and which affects our action because the agent accepts it as a reason. And reasons, once again, can be reduced to beliefs, together with those other considerations which do not qualify as beliefs simply because there is no question of the agent's failing to realize that they are so.

Nevertheless, the claim that reasons presuppose wants will now have to be replaced by the claim that reasons require wants if they are to be effective in action. But if the earlier account trivialized Davidson, this present account trivializes Hume. For if to want to do something is to have some reason which actually inclines us to act, then to say that reasons lead to action only when we also want to perform those actions is not to say that some extra element needs to be present before the agent will act on his reasons. It is to say only that he will act on his reasons only if those reasons are effective. In short, reasons will be effective only when they are effective.

### III

One major merit of this account of wanting is the light it throws on the vexed questions of whether wants can cause actions. It is often argued that Hume established once and for all that causes and effects must be logically independent. But since wanting and acting are logically connected, wants therefore cannot be "Humean" causes of action. Similar arguments are offered for intentions, purposes, motives, volitions, etc.—though not, significantly, for reasons.

The standard form of this argument is that since we describe or identify a want in terms which we use also to describe or identify the action, there is therefore a conceptual link between wanting to do it and doing it.<sup>3</sup> But a conceptual connection of this sort—if it is a conceptual connection, and not merely a terminological one—in no way conflicts with a causal connection. The fact that we call the particular want "the want to do *x*" no more shows that it cannot be causally related to my doing *x*, than the fact that we call a certain smell "the smell of coffee" shows that it cannot be causally related to coffee. What needs to be shown, if it is to be argued that wants cannot be "Humean"

<sup>3</sup> Cf. A. I. Melden, *Free Action* (London and New York, 1961); T. F. Daveney, "Intentions and Causes," *Analysis*, vol. 27 (1966-67), pp. 23-28.

causes of actions, is that it is logically impossible for the particular want not to result in the particular action. The mere fact that a want is described or identified or even defined by reference to some action no more shows that the action must, with logical necessity, follow upon that want than the fact that the smell is described, identified, even defined, by reference to coffee shows that coffee must, with logical necessity, be present whenever the smell is. Coffee and the smell of coffee are, in Hume's terms, distinct existences: one can occur without the other and that, it seems, is all that is required for "Humean" causation.

What is needed, then, is an argument that it is a conceptual truth that the want will lead to the action, and this we find in the argument<sup>4</sup> that if a man has the chance to do  $x$ , has no countervailing wants, and yet does not do  $x$ , then it follows necessarily that he did not want to do  $x$ . Part, at least, of what we mean by saying that he wants to is that he will, other things being equal. It is, therefore, a conceptual truth, a logical necessity, that if a man wants to perform some action  $x$  then, other things equal, and in particular given no countervailing wants, he will do  $x$ . Yet those who make this point do not conclude, without reservation, that wants do not cause actions.<sup>5</sup> For there are many other, perfectly acceptable cases of causation where cause and effect are linked in just this way. It is, for example, a conceptual, and not merely contingent, fact that taking poison leads, other things equal, to sickness and even death, or that scars are the results of wounds, but this does not conflict with the obvious fact that poisons cause sickness or that scars are caused by wounds.

Does this mean that Hume was simply mistaken in asserting that events which are logically linked cannot be causally related? Not at all, for Hume nowhere makes that claim. For a start he does not use this contemporary jargon; and in any case his whole point is that logical or conceptual links hold only between ideas, not between existences or occurrences. There is and can be no logical link between the events of taking poison and feeling ill; the connection holds only between the terms in which we describe those events. To argue that because two terms, such as "wanting" and "acting," are logically related the existences or

occurrences which they designate, wanting and acting, must also be logically related, and hence cannot be causally related, is to make just the mistake Hume was pointing out, the mistake of reading a conceptual connection between our ideas onto the things themselves, where no conceptual connections can be. Yet this is the argument put forward in his name! On the other hand it would be foolhardy to argue that there can be no logical links between terms which describe or designate causes and effects, for there is just such a link between the terms "cause" and "effect" themselves.

What Hume does claim, of course, is that causal connections cannot be established a priori: "if we reason a priori, anything may appear able to produce anything." But the fact that it is a conceptual truth that poisons cause sickness does not affect the claim that "it is only experience . . . which enables us to infer the existence of one object from that of another." For it is only from experience that we discover that various things are poisons. It is a conceptual truth that poisons cause sickness because to classify something as a poison is already to classify it as something which causes sickness, but it remains a contingent, empirical matter whether this, that, or the other substance is a poison, does cause sickness. And similarly for wounds and scars.

So the position seems to be that while it may be a conceptual truth that wants lead to action, this need not conflict with the claim that wants cause actions, so long as to classify something as a want is already to classify it as something which leads to action, where it remains a contingent, empirical matter whether this, that, or the other thing is a want, does lead to action. But at this point there is a further difficulty. For while those things which are poisons, such as arsenic or strychnine, can be identified and characterized independently of their tendency to produce sickness, is it possible to identify and characterize wants, other than as tendencies or inclinations to act in the appropriate ways? What, after all, is a want—especially when "want" is interpreted as broadly and weakly as we have interpreted it—except a disposition of the agent to act in a certain way, given the chance and no countervailing wants? "The want is just the likelihood that a certain sort of action will

<sup>4</sup> Charles Taylor, *The Explanation of Behaviour* (London and New York, 1964), p. 33; Norman Malcolm, "The Conceivability of Mechanism," *The Philosophical Review*, vol. 77 (1968), p. 49.

<sup>5</sup> Norman Malcolm, *op. cit.*, p. 59; Charles Taylor, "Explaining Action," *Inquiry*, vol. 13 (1970), Sect. I pp. 54-59.

take place, so it accounts for the action only verbally."<sup>6</sup> And despite the most subtle enquiries<sup>7</sup> it proves enormously difficult to give the want any substance, any existence, except as a tendency to act in the relevant way. This, it seems, and not any logical link between want and action, is what stands in the way of our regarding wants as straightforward causes of action. It isn't that wants and actions are logically connected, for logical links can hold only between terms or ideas, not existences; it is rather that wants and actions appear not to be distinct existences which might be connected by constant conjunction in the way that Hume describes.

There is by now an enormous literature devoted to these topics, but if our earlier analysis of wanting is anywhere near correct, the entire controversy rests on a misunderstanding. Wants, we said, are effective reasons, or reasons which are effective as reasons; to want to do something is to have some, so far unspecified, reason to do it, a reason which would actually be effective in your doing it, given the opportunity and no countervailing wants. It is therefore an error to think that the cause of an action will include wants as well as beliefs,<sup>8</sup> analogous to thinking that any cause must include causation, or a power to bring about effects, as well as the particular factor which is the cause. Causation is not part of the cause; it consists simply in the fact that the factor in question is a cause, does have effects. Similarly my wanting to do something is not part of the cause of my doing it but consists simply in the fact that some reason does cause, or lead, me to do it. It is the causing, not the cause. This is why wants are unfavorable candidates as causes of action; this is why wants, as here conceived, amount to no more than tendencies to act in the appropriate ways; but it is also why this in no way shows that actions done because the agent wants to do them lack causes altogether. The cause will be those reasons, standardly beliefs, which lead the agent to act in that way; his wanting to consists simply and solely in the fact that some such beliefs or reasons affect, or at least influence, his behavior.

The failure to recognize that beliefs, by themselves, can be causally sufficient for action is in large part the legacy of Hume's artificial distinc-

tion between impotent reason and overmastering passions. Yet it is surely obvious that various beliefs—that the house is on fire, that there is a snake at your feet, that something ought to be done—can and do move people to action: they act because they have those beliefs; if they did not have those beliefs they would not act as they do. It is, of course, puzzling and problematic just how beliefs move people to action, but this is no more a *philosophical* question than the question of how flames ignite gases or why apples fall down rather than up. There is also the problem of why certain beliefs move some people to action but not others, or move some people in one way and others in another, and it may seem that we need to invoke varying wants to account for these differences. But we have already seen that the appeal to wants, in our formal sense, does not explain this difference; it merely repeats the point that some people are motivated by some considerations and others by others. The question of why this should be so is, presumably, a question for empirical psychology. Certainly it too is hardly a philosophical issue, no more than the question of why certain chemicals kill some insects but not others.

#### IV

We can now return to the question of the differences between reasons and (other) causes. Clearly there are differences: whatever else they may be, reasons are not like physical forces. But the crucial issue is whether these differences show that actions done for reasons differ in some important and essential respect from events in the world of inanimate nature. There is space here to consider only the two most important suggested areas of difference.

There is first the familiar point that an agent shapes his actions in terms of how he sees and understands things. This understanding may in fact be mistaken, but if we are to understand his actions we must understand what he understands. Putting this in our terms, a man acts as he does because he believes what he does; the fact that these beliefs may be false does not prevent them from explaining his action. But surely, it may be said, something which is not so can never be a

<sup>6</sup> C. H. Whiteley, "Mental Causes," in G. N. A. Vesey (ed.), *The Human Agent*, Royal Institute of Philosophy Lectures 1966-67 (London, 1968), p. 107.

<sup>7</sup> E.g., D. Pears, "Desires as Causes," in G. N. A. Vesey (ed.), *op. cit.*, pp. 83-97.

<sup>8</sup> Cf. Donald Davidson, "Actions, Reasons, and Causes," p. 686; A. Goldman, *A Theory of Human Action* (Englewood Cliffs, N.J., 1970), ch. 3.

cause, and hence reasons will not be causes. More than that, an agent's reason is, strictly, the content of a belief, what he believes, and it is hard to see how a belief-content—whatever that might be—could be a cause of anything. The connection between reasons and actions will be rational, or logical in a broad sense, not empirical or physical.

This sort of point can be made in a variety of slightly different ways: Reasons are not events or circumstances which produce or bring about results; they are intellectual considerations which make behavior intelligible or appropriate. Reasons justify, causes explain. Reasons can be assessed as good or bad, or even as no reason at all; causes either cause or they do not. Causes connect with their effects via relations of regularity; reasons connect with actions via relations of relevance. Agents know their reasons for doing what they do "without observation," and certainly without any appeal to empirical uniformities, but causes are discovered by experiment and observation. Events brought about by causes are law-governed; actions done for reasons are rule-following. And so on.

Our theory of reasons explains these differences: It is precisely because reasons are beliefs that questions of justification and assessment and relevance and rationality and intelligibility and the like arise for them in a way that they do not arise for (other) causes. Moreover these differences do rest on a difference in kind between reasons and causes. For as we saw earlier, it is the content of the agent's belief, *what* he believes, which constitutes his reason, the agent-reason—and which we assess as good, bad, or no reason at all. But it is the existence of this belief, *that* he believes it, which provides the explanation, and will hence be the cause, if anything is. Nevertheless this does not mean that actions done for reasons are not also the results of causes. If anything it implies the very opposite, since where the thing believed provides a reason, the believing of it will provide a cause. The only exception to this will be in cases of rationalization, as in our example of post-hypnotic suggestion, where the belief which constitutes the agent-reason does not provide the reason, and is not the cause of the action. Indeed reasons will never provide causes if, as some maintain, all our reasons are simply rationalizations. But this is hardly what people have had in mind when they have denied that reasons are causes! For the main interest in the distinction has stemmed

from the hope that it will enable us to demonstrate that human actions done for reasons exhibit a freedom which events brought about by causes do not.

This brings us to a second major suggested point of difference between reasons and (other) causes, one that is probably best summed up in Leibniz' dictum that "reasons incline without necessitating." One way of putting the traditional problem of free will is in the form of a dilemma: if on the one hand our actions have causes, then no one could ever do other than what he does; but if on the other hand our actions lack causes, then they will be chance, random, inexplicable, scarcely actions at all. And one way of avoiding this dilemma is to argue that although actions lack causes this does not make them chance, random, or inexplicable, since they can still be accounted for in terms of reasons. Moreover, this explanation will not carry the implication that the agent could not have done otherwise, precisely because reasons incline without necessitating.

Now this claim that reasons incline without necessitating can easily be made tautologous. For we can interpret "incline" as meaning, roughly, "gives reason to," in that to say that his dishonesty inclines me to sack him is to say that his dishonesty is, for me, a reason for sacking him, and to say that I am inclined to forgive him is to say I find some reason to forgive him. "Reasons incline," on this interpretation, means no more than that reasons are reasons. Similarly, the claim that reasons do not necessitate seems to amount to the fact that where a man acts for a reason, the reason which explains his action does not force him to act as he does, making him act willy-nilly, whether he wants to or not. And it seems to follow from what we have said about reasons and wants that where a man acts for a reason then, on one interpretation at any rate, he does what he most wants to do. An agent may do something he does not want to do, yet do it for a reason; he may even do something which, he well realizes, he has more or better reason not to do; but if he acts for a reason then, necessarily, he will do that which his reasons most incline him to do, and hence, in that sense, he will do what he most wants to do. It is in this sense that reasons do not necessitate, do not make us act whether we want to or not: if you act for a reason it cannot be that you would do it even if you had reasons influencing you more strongly toward some other course of action.

But causes, it will be said, do not incline, they necessitate. What does this difference amount to? Certainly there is no idiom pertaining to causes which corresponds to the phrase "reasons incline," but this may only be because causes do not qualify as causes unless they actually produce an effect, while a reason will be a reason even if it is outweighed by other reasons, and so does not result in action. To say that reasons incline is to say that if *A* has a reason for doing *x* then *A* will do *x*, other things being equal, given no obstacles or countervailing factors. So might we not introduce the notion of, say, a "potential cause," where a potential cause is something which will be followed by the effect, other things being equal, given no obstacles or countervailing factors? "Potential causes" will then incline in just the sense that reasons do: a dose of poison, for example, will be a potential cause of death, in that it will result in death unless an antidote or emetic is administered. Indeed physical forces, one of the more obvious varieties of cause, are potential causes in just this sense. It does not follow just from the fact that a particular physical force is exerted on some item that it will move appropriately. That follows only given that other things are equal, in particular that there are no other forces exerted on the same item which keep it in its place. There seems, then, no reason to deny that physical forces incline in just the way that reasons do.

Yet we do not say it; we do not speak of the dose of poison as inclining me to die, or of the impact of the cue as inclining the billiard ball to move, as we do speak of his dishonesty inclining me to sack him. The difference lies, I think, in a fact we noted earlier, that a reason is not a reason for the agent, not one of his reasons, unless he accepts it as such. Poisons and physical forces will affect me whatever I think of them, indeed whether I know of their existence or not, but a reason will be a reason for me only insofar as I regard it as a reason. To put it metaphorically, a belief requires my consent to become a reason for me, but causes are causes no matter what I think.

It is in this sense that what a man does, when he acts for a reason, is up to him: it is dependent on his decisions, attitudes, and opinions, and hence, to that extent, his responsibility. I accepted that reason as reason enough to act in that way; if I had not made that judgment, been of that opinion, I would not have acted as I did; and that is why my action is something I may be held accountable for. But while this is an important difference

between reasons and (other) causes, it is not sufficient to support any strong claim for human freedom. For it may well be that the further factors which explain why an agent accepts the reasons that he does are factors which are in no sense up to him, or dependent upon his decisions, attitudes, or opinions. In that case what a man does, even where he acts for reasons, is up to him only in a fairly restricted sense. It will still be possible to argue that ultimate responsibility lies not with him but with the circumstances which make him the man that he is, a man who acts for reasons like these.

What, then, of the other suggested difference, that whereas reasons merely incline, causes necessitate? It is clear that some causes do necessitate in precisely the sense that reasons do not, that they make us do things willy-nilly, whether we want to do them or not: if I am struck in the small of the back by a passing bus, I will move forward irrespective of my reasons and wants. Nevertheless this difference does not hold between reasons and (other) causes in general, for not all causes necessitate their effects in this current sense. The wind causes the branches to move, but it does not make them move whether they want to or not; the notion of necessitation, explained as we have explained it, can apply to this case only on the curious misapprehension that branches might have beliefs about what to do. Thus the familiar claim that actions done for reasons are free inasmuch as they are not necessitated in the present sense carries with it the unwelcome consequence that the branches are no less free than I am—or that I am no more free than they.

So if it is to be claimed, quite generally, that all causes necessitate in some way which reasons do not, the word "necessitate" will need some other interpretation. On the present interpretation that claim is, at best, a metaphor. The sense in which all causes, qua causes, do necessitate is, presumably, that in which *A* necessitates *B* to the extent that if *A* happens *B* must happen, in that it is empirically impossible for *A* to happen and *B* not to happen: the wind necessitates the movement of the branches in that it is empirically impossible for the wind to blow as it does and the branches to remain still. Now it may seem that reasons do not necessitate in this sense either, since where an agent acts for reasons it is possible for him to have just those reasons and yet not act in that way. The case of *akrasia* is just such a case, where the agent agrees he has more reason to do something

than not to do it, and yet he does not do it.<sup>9</sup> But in just the same way it is perfectly possible for the wind to blow as it does and yet not move the branches—because we have meantime fixed them to stop them moving. The point is simply that the force of the wind—or, similarly, the agent's reason—is not *by itself* enough to make any other outcome impossible. What is impossible, apparently, is that the wind blow and the branches remain still, given that everything else remains as it is, that the circumstances stay exactly the same. But perhaps this is equally true when agents act for reasons; perhaps it is empirically impossible for an agent to have just those reasons and yet act differently, even though everything else about him and his situation remains exactly as it was. I do not know whether this is so or not. In any case it is not something to be settled by an analysis of reasons and acting for reasons.

There is, however, one consideration to suggest that reasons do necessitate in the present sense, just as (other) causes do. For if an agent's behavior is to be explicable at all, there will presumably be some factor, independent of the various beliefs which constitute his reasons, whose presence in the one case explains why he acts for those reasons, and whose absence in the other case explains why he does not act, despite having those reasons. So

to insist that where an agent acts for certain reasons it was nevertheless possible for him to have just those reasons and yet not act as he did, even in those self-same circumstances, is to insist that there is no such further factor which explains why he acts, if he does act, or why he doesn't act, if he doesn't act. Consequently what he does, whether he acts or not, is in a very real sense inexplicable. Certainly it lacks a sufficient reason, in Leibniz' sense of something which explains why things are so and not otherwise.

Thus the claim that reasons do not necessitate as causes do has the somewhat unpalatable consequence that when agents act for reasons it is, in the end, impossible to explain why they should act for those reasons, rather than ignoring them. Some libertarians are prepared to embrace this paradox, in the interests of human freedom and dignity. But one thing is clear: this is no way out of the original dilemma, which argued that our actions are either caused or inexplicable. I suspect that it is the realization of this point—that reasons either constitute sufficient conditions, i.e., conditions which are sufficient in the circumstances, or they do not adequately explain—which has led many<sup>10</sup> who originally insisted that reasons are different from causes to conclude, in the end, that reasons are but one variety of cause.<sup>11</sup>

University of Warwick

Received May 14, 1973

<sup>9</sup> Cf. Irving Thalberg, *Enigmas of Agency* (London and New York, 1972), p. 29.

<sup>10</sup> Contrast A. C. MacIntyre, "Determinism," *Mind*, vol. 66 (1957), pp. 28–41, with A. C. MacIntyre, "The Idea of a Social Science," *Proceedings of the Aristotelean Society*, Supplementary Volume XLI (1967), p. 100.

<sup>11</sup> This theory of reasons and wants was originally put to me by Professor Phillip Griffiths (see his "Reasons and Causes" in *Education and the Development of Reason*, ed. by R. H. Dearden, P. H. Hirst, and R. S. Peters [London, 1972].) Much of what I say could be regarded as an exposition of his views, except that he has since moved on to another position. I am enormously indebted to him, but he is responsible only causally for what I say here. I have also had the benefit of discussions at several universities in the United States. Among the many people who have made helpful comments I would like particularly to mention Haskell Fain, Nancy Holmstrom, and Laird Addis.



### III. THE ONTOLOGICAL ARGUMENT AND THE CONCEPT OF SUBSTANCE

J. MICHAEL YOUNG

#### I

SEVERAL years ago Professor Norman Malcolm pointed out that there seem to be two distinct arguments in Anselm's *Proslogium*—one in Sect. II, the other in Sect. III—either of which might plausibly be identified as the ontological argument.<sup>1</sup> Malcolm rejected the first of these arguments, but he tried to defend the second. His defense was unsatisfactory, I think, and the reasons for this were pointed out at the time by Alvin Plantinga.<sup>2</sup> What was not pointed out, however, is that Malcolm may have been fundamentally wrong in his interpretation of Anselm.

On Malcolm's interpretation, the argument of *Proslogium* II is an attempt to prove that God exists. The argument of *Proslogium* III, he thinks, is more powerful; for it purports to show that it is necessarily true that God exists (whence it would follow, of course, that God exists). On Malcolm's view, then, the second argument is, as it were, a second and more powerful shot at the same target. But Anselm's text seems to read a bit differently. In the first place, Anselm doesn't claim to show in the first argument that God exists. He claims to show, instead, that we cannot doubt God's existence without lapsing into absurdity. (This distinction may seem fine, but I think it will turn out to be important.<sup>3</sup>) In the second place, Anselm does not seem to regard his second argument as an alternative—not even as a more powerful alternative—to the first. In fact, as he develops the second argument he seems rather to take for granted the conclusion of the first. For what he argues is that this God, whose existence (by the first argument) we cannot doubt, cannot be conceived not to exist.

It seems, then, that Anselm's two arguments

are meant not as alternative routes toward the same conclusion, but as two interconnected parts of a single argument. The aim of the first part is to show that we cannot intelligibly doubt that God exists. The aim of the second part, as I read it, is to determine the *mode* of God's existence. It is to show that God, whose existence (by the first part of the argument) we cannot doubt, exists necessarily—that he exists, not by virtue of a dependency on anything else, nor merely accidentally, but simply by virtue of his own nature.<sup>4</sup> (This is what Anselm means, I think, when he says that God "cannot be conceived not to exist.")

I do not want to argue about the proper exegesis of Anselm's text, however. Indeed, even if I am right about how the text is to be interpreted, I do not want to defend or attack Anselm's version of the argument. I *do* want to argue, however, that if we interpret the ontological argument in the way I have just suggested—as having this two-part, compound form—then a defensible version of it can be constructed. In fact, I think that such a version lies ready to hand in Part I of Spinoza's *Ethics*.

I will try to show that if Spinoza is correct in his theory of substance, if he is also correct about a basic principle concerning causality, and if he is correct in thinking that substance has to be conceived as God, then it follows that we cannot intelligibly doubt the existence of God, and that God exists necessarily. Putting it more roughly, but more plainly, I want to show that if we substitute Spinoza's concept of substance for Anselm's notion of a being "than which nothing greater can be conceived," we get a defensible version of the ontological argument. (It might even be argued that this is an interpretation of Anselm,

<sup>1</sup> "Anselm's Ontological Arguments," *The Philosophical Review*, vol. 69 (1960), pp. 41–62; reprinted in *The Ontological Argument*, ed. by Alvin Plantinga (Garden City, 1965), pp. 136–159.

<sup>2</sup> "A Valid Ontological Argument?," *The Philosophical Review*, vol. 70 (1961), pp. 93–101; also reprinted in *The Ontological Argument*, *ibid.*

<sup>3</sup> See below, Sect. III, and Sect. IV, comment (4).

<sup>4</sup> Henceforth when I say that something "exists necessarily," this is what I shall mean. To say that God "exists necessarily," then, is not to say that it is necessarily true that God exists. It is rather to make a claim about the mode of God's existence.

and not a substitution; for it might be argued that Spinoza's concept of substance serves simply to clarify Anselm's rather intuitive concept of "that, than which nothing greater can be conceived." But again, I shall leave the exegetical question aside.)

I propose, then, that we understand the ontological argument as consisting of two complementary parts. In the next section I shall discuss Spinoza's theory of substance and apply it to the *second* part of the ontological argument, showing why, given Spinoza's view, God exists necessarily. In the third section, I shall return to the *first* part of the ontological argument, showing why, given Spinoza's view, the existence of God cannot be intelligibly doubted. In the last section I shall add a few concluding comments.

## II

In this section I intend to discuss Spinoza's theory of substance, which he develops in the first seven propositions of his *Ethics*. In doing so, I intend to ignore any aspect of his position which derives from his specific view that there is but one substance, namely, God, or "Being absolutely infinite." I intend, that is, to focus on his theory of substance, while ignoring his specific view as to how substance is to be conceived. It is worth noting, I think, that this will not involve any distortion of Spinoza's view. He himself carefully separates his abstract treatment of substance from his specific views concerning the nature of substance. He first argues that substance exists necessarily (*Ethics*, I, vii)<sup>5</sup>. Only after he has shown this does he introduce the concept of God, or "Being absolutely infinite," into his proofs. Moreover, his demonstration that God exists necessarily rests squarely on his treatment of substance, for he simply argues (*Ethics*, I, xi) that God is substance, and substance (as already demonstrated) exists necessarily.

The arrangement of Spinoza's demonstrations is no accident, either. God exists necessarily, on Spinoza's view, simply because he is substance, and not because he is substance of this or that specific nature. By focusing on this concept, accordingly, we shall direct our attention to the heart of the ontological argument. Indeed, there is an even stronger point. By focusing on the

concept of substance we enable ourselves to realize that the ontological argument need not be thought of as an argument dealing with God! The argument deals simply with substance; and it is a further question what the specific nature of substance is. As I shall indicate in passing, we can easily imagine a metaphysical view which would employ the concept of substance, but which would not conceive of substance as God. And a proponent of such a view would also be committed to the conclusion that substance, as he conceives it, exists necessarily.

Spinoza delineates the main features of substance by means of the distinctions (1) between that which is in itself and that which is in another, and (2) between that which is conceived through itself and that which is conceived through another (*Ethics*, I, Df. iii). The first of these distinctions serves to indicate that substance is metaphysically fundamental or basic. To say that something is "in another" is to say that the thing in question is simply a modification or determination of something else, simply a way—or as Spinoza terms it, a *mode*—in which something else exists. Spinoza defines substance as that which is not in another, but rather "in itself." He means, I take it, that substance is something of which other things are modes, while it is not itself a mode of anything else. The second distinction seems, at least initially, to mark a somewhat broader point about substance. He says that substance, besides being in itself, is also conceived through itself—that it is something "the conception of which does not need the conception of another thing from which it must be formed" (*Ethics*, I, Df. iii). To say that something is "conceived through another" is to say that in order to conceive it properly, in order to understand what it is, we need to appeal to the concept of something other than it itself. Substance, however, is not conceived through another. It is something which is not to be understood in terms of anything else, but which instead is that in terms of which we understand other things.

It may be helpful if we "illustrate" Spinoza's concept of substance by sketching some rough outlines of a metaphysical view which is much simpler than his, but which nevertheless employs his concept of substance. Let us imagine a form of materialism, in which it is matter, and not God

<sup>5</sup> *Spinoza Selections*, ed. by J. D. Wild (New York, 1930), p. 98. The translation of the *Ethics* is by W. H. White. I follow the usual practice: "*Ethics*, I, vii" refers to proposition vii in part I of the *Ethics*.

or "Being absolutely infinite," which is taken to be substance.

According to a materialist of the sort imagined, particular objects, such as pebbles or men, exist simply as determinations of matter. It is when matter is so determined as to constitute a multiplicity of elementary components, and these are arranged in such a way as to constitute components of a higher order, and these in turn are arranged in such a way as to constitute components of a still higher order, and so on through many levels, that objects of this sort exist. Indeed, such objects are nothing but matter existing with such complex determinations. And for a materialist of this sort, all such objects are thus "in another," since they are simply finite (albeit complex) modes in which matter exists. Matter itself, however, would not be "in another." For what makes the materialist a materialist is precisely his contention that matter is metaphysically fundamental. He holds, then, that matter is "in itself." He also holds that matter is "conceived through itself." For he maintains that we understand what a thing is only when we understand it as a complex determination of matter. And he maintains that matter itself cannot be understood in terms of anything more basic than itself.

Thus the materialist that we have imagined employs Spinoza's concept of substance, even though he embeds it in a different metaphysical theory. Given that he does so, moreover, he commits himself to Spinoza's version of the second part of the ontological argument, as we shall see.

Put briefly, Spinoza's version of the argument runs as follows:<sup>6</sup> Everything is either caused by another, or cause of itself. That is to say, things are either dependent upon other things for their existence, or they are not thus dependent, but exist just by their own natures, and hence necessarily. Substance cannot be caused by another, however. If it were, it would have to be conceived through that other thing, since the cognition of an effect depends upon and involves the cognition of its cause (*Ethics*, I, Ax. iv). That would be absurd, however, since substance is, by definition, conceived through itself. Substance, therefore, cannot be caused by another. It must rather be cause of itself, i.e., it must exist necessarily.

Now it seems to me that given Spinoza's principles, including his conception of substance, his

argument is basically correct. This is hardly apparent, though, when the argument is presented in this very brief form. For it is not at all clear how the distinction between that which is caused by another and that which is cause of itself comes to be connected with Spinoza's other two basic distinctions, and hence with the concept of substance. Nor is it evident that Spinoza's *reductio* will actually hold up—that when we reject the possibility of substance being caused by something else, we have to affirm that it exists necessarily. Let me try to remedy these two defects.

Consider first the connections between Spinoza's three basic distinctions. He sets it down as a principle (*Ethics*, I, Ax. iv) that the cognition of an effect depends upon and involves the cognition of its cause, so that that which is caused by another is also conceived through another. The force of this principle seems, moreover, to be twofold. He seems to mean that we cannot conceive of an effect properly unless we conceive it as the effect of its cause. He also seems to mean that the concept of the cause must enable us to understand why the effect follows. Ultimately, though, both points come to the same thing—namely, that the connection between a cause and its effect cannot, on Spinoza's view, be merely external. It must rather be implied in the very concept of the cause and the effect. The concept of the effect must imply that it is dependent on its cause, and the concept of the cause must imply that its effect follows from it.

It would of course take a lengthy argument to defend Spinoza's principle. For our purposes, however, such an argument is not necessary. What we want to know is whether, given his conception of substance and his principle concerning causality, Spinoza can rightly conclude that substance exists necessarily (and, as I shall argue in the next section, that its existence cannot be doubted). What the critics of the ontological argument have traditionally argued, after all, is that this inference is not legitimate. Let us grant Spinoza his principle, then, and see whether his conclusion follows.

It follows from Spinoza's principle that what is caused by another is also conceived through another. It is also the case, however, that what is conceived through another is in another. For to say that something is conceived through another is to say that we can only understand what the thing is by conceiving it in terms of something else.

<sup>6</sup> This brief version follows the lines suggested in *Ethics*, I, vi, cor. As I comment further my own version will come closer to the full argument of the first seven propositions taken together.

It is to say that the very being of the thing in question is constituted by something other than itself, and hence that it is in another. What is caused by another is therefore a mode, in Spinoza's terms. Moreover, the converse also holds: modes are always caused by another. For a mode, something that is in another and conceived through another, has its existence constituted by something other than itself, and thus it is caused by another. Indeed, the simplest thing to say is just that the distinction between being a mode and being caused by another—the distinction between causal dependency and what we might call metaphysical dependency—ultimately vanishes on Spinoza's view.<sup>7</sup>

Now the point of importance in all this is that substance, which is not in another, cannot be caused by another. To think of substance as caused, we would have to think of it as in another and conceived through another. We would have to think of it, that is, as a mode of something else; and obviously, that would be self-contradictory. (If the materialist that we imagined earlier were to suppose that matter is dependent upon something other than itself for its existence, he would contradict his own materialism; for he would thereby contradict his own view that matter is substance. And a similar point holds, of course, for any metaphysician who employs the concept of substance.)

This is the point, then, that Spinoza develops in the first six propositions of the *Ethics*, where he works out the connections between his three basic distinctions. And this is the point that he makes use of in his demonstration of proposition vii. Substance cannot be caused by another, he argues. It must, therefore, be cause of itself. It must exist, that is to say, not because of anything else, but simply of its own nature, i.e., necessarily. As he puts it: "It pertains to the nature of substance to exist."

Spinoza's argument invites an obvious reply. Might it not be the case that substance is neither caused by another nor cause of itself, but simply uncaused—that it exists neither as an effect of something else nor of its own nature, but simply as a matter of brute, unintelligible fact? Unfortunately, Spinoza does not attempt to meet this objection. He seems simply to take it for granted that the "third alternative" is merely illusory. It is not difficult, however, to see how he might respond to this sort of criticism.

When Spinoza asserts that substance is cause of itself, he does *not* mean to say that substance is related to itself in the same way that it is related to its modes. That would obviously be absurd. What he *does* mean to say, I think, is that substance exists of its own nature, necessarily, and that it could not be otherwise. And it is just this ascription of necessity that the critic means to call into question. He means to suggest that it might be the case (so far as Spinoza's argument has shown) that substance exists, not necessarily, but just as a matter of accident, or of brute fact. This suggestion would have to be dismissed as absurd, however, so long as we accept Spinoza's conception of substance. The critic means to suggest that while substance may exist as a matter of fact, this is *merely* a matter of fact. He means to suggest, therefore, that things might well have been otherwise. What this amounts to, however, is a suggestion that substance might have been other than substance! For on Spinoza's view, substance is that which is metaphysically fundamental. To consider the possibility that something might not exist is therefore to consider the possibility that substance might be determined so as not to constitute the thing in question. And hence to consider the possibility that substance might not have existed is to consider the possibility that substance might have been determined in such a way that substance would not have existed. But that, of course, is absurd. We can no more conceive the possibility of substance being other than substance than we can conceive of the number three being other than the number three.

At this point the critic might begin to question the intelligibility of Spinoza's concept of substance. Again, however, I am not concerned with questions of this sort. My aim is simply to determine whether, making use of Spinoza's concept of substance, a defensible version of the ontological argument can be constructed.

As far as the *second* part of the argument is concerned, I think that the answer is "Yes." Spinoza has shown in the first seven propositions of the *Ethics* that given his concept of substance, and given the relevant assumption concerning causality, it follows that substance exists necessarily. Hence, if we go on to assume that Spinoza is right in thinking that substance should be understood more specifically as God, it follows that God

<sup>7</sup> This is vastly oversimplified, for it ignores the distinction between finite and infinite modes. But even when this distinction is taken into account, it remains true that to be an effect is to be a mode, and conversely.

exists necessarily, and not accidentally or by virtue of a dependence on something else.

As far as the *first* part of the ontological argument is concerned, however, the answer is not yet clear. In the first seven propositions of the *Ethics* Spinoza speaks—as I have spoken in discussing his views—as though there were no doubt about whether there is substance, but only about the mode of its existence—whether it exists necessarily, or in some other way. It seems perfectly reasonable, however, to doubt whether there actually is any substance with this mode of existence.

As far as I can see, Spinoza does not address himself to this type of doubt at any point in the *Ethics*. I think that it is possible to show, however, that given his theory of substance, it follows that we cannot intelligibly doubt the existence of substance. To this task I now turn.

### III

It is this first part of the ontological argument which has been the chief source of controversy. It seems, in fact, that discussion of the ontological argument has focused almost exclusively on this first part. In this section I shall try to show that a defensible version of this first part of the argument can be constructed, utilizing Spinoza's concept of substance. When I do so, however, I shall be developing the argument in a way which differs, I think, from the usual way of developing it. This requires some preparation. Accordingly, I shall begin by describing what seems to be the standard way of developing this first part of the ontological argument. Then I shall argue that the argument, developed in this standard way, is incorrect—although it is no simple matter to show just why. And finally, I shall suggest that the argument might be developed differently, in a way that is basically correct.

To begin with, then, there is a standard way in which *both* the critics and the proponents of the first part of the argument usually develop it. On this standard version, the argument is presented as an attempt to show that the judgment that God exists is analytic. It is presented, that is, as an attempt to show that existence is somehow implicitly contained in our very concept of God, so that any denial of his existence is self-contradictory, and in this sense unintelligible.

Now if we take this as the proper way of developing the argument, then I think that we cannot help but regard the argument with suspicion. It is hard, in the first place, to believe that such a momentous conclusion could be established so easily. More seriously, it is hard to believe that an argument of this sort could ever establish the existence of anything; for it is hard to believe that we could ever contradict ourselves simply by denying that there is something corresponding to an idea we have. Our intuitive doubts about the argument are, I think, basically correct. As it turns out, however, it is no simple task to show just why the argument fails—to convert our suspicions into a refutation. Certain difficulties arise when we try to state just what the error in the argument is. And these difficulties turn out to be instructive.

It is tempting to suppose that we might reject the argument (as it is normally interpreted) by means of a direct frontal attack. It is tempting, that is, to suppose that we might grant the major claim—the claim that existence is somehow contained in our idea of God—without thereby being compelled to grant that God actually does exist. After all, we are tempted to argue, it may be true that in conceiving of something as a body we necessarily conceive it as extended, but this obviously has no bearing on whether there actually is anything that is a body. By analogy, then, even if it is true that our concept of God requires us to conceive of him as existing, it does not follow that there actually is a God of the sort that we conceive.

Tempting as it is, however, this line of attack simply fails. If we grant that existence is somehow contained in our concept of God, then we are granting that a denial of God's existence is self-contradictory, and in this sense unintelligible. But if so, then it makes no sense to regard it as a hypothetical matter whether there is anything that falls under the concept of God—whether, that is, God exists. And the purported analogy between the judgment that God exists and other analytic judgments therefore breaks down.<sup>8</sup>

If we are to reject the argument, then, it will have to be on stronger grounds than these. What we shall need to argue, it seems, is that the major claim—the claim that existence is somehow contained in our concept of God—is simply wrong, and that the judgment that God exists cannot be analytic. And this is just what the critics of this

<sup>8</sup> Descartes makes just this point in ch. V of the *Meditations*.

first part of the ontological argument have typically argued. In fact, just as there seems to be a standard way of developing this first part of the argument, so too there seems to be a standard way of criticizing it. The criticism is expressed in a variety of forms and idioms, but the thrust is basically this: When we assert that something exists, the predicate of our judgment does not merely indicate that we conceive of the thing in a certain way, as having a certain determination or property. It serves instead to indicate that there is something of just the sort that we conceive—something that has the determinations or properties thought in the subject concept. Clearly, then, the judgment that God exists cannot be simply an explication of our concept of God. Clearly, in other words, it cannot be analytic. And therefore the argument fails.

In assessing this standard line of criticism, what is important is that we examine the basis on which it is made to rest. The critics do not merely *assert* that the judgment that God exists cannot be analytic. They claim to *show* that it cannot be analytic by appealing to a general view concerning the difference between existence and other predicates. They argue that, as Kant puts it, existence is not a “real” or “determining” predicate—that when we assert that something exists, we do not assert anything about how the object is conceived. Rather, we simply “posit the subject in itself with all its predicates, and posit it as an *object* that stands in relation to our *concept*.”<sup>9</sup> From this it follows, they argue, that a judgment in which we assert that something exists cannot be analytic, and in particular, that the judgment that God exists cannot be analytic.

It appears that this line of objection (developed in various ways, and expressed in various idioms) is widely accepted. I doubt, however, whether it is really as formidable as its widespread acceptance would seem to indicate. In fact, if I may turn one of Kant’s famous phrases back against him, it seems to me that this standard line of objection is nothing more than a miserable tautology—either that, or it is merely a *non sequitur*.

If the critics mean to argue that we cannot regard the judgment that God exists as *merely* an assertion about how we *conceive* of God, then their “argument” is really nothing but a tautology. For this is merely to repeat the basic point that

when we predicate existence, we mean to posit an object over against our concept. It is, moreover, a miserable tautology, for this basic point is not the least bit controversial! For obviously, proponents of the ontological argument have not meant to say that the judgment that God exists is *merely* an assertion about our *concept* of God. They have interpreted the judgment, as Kant himself would, as a positing of an object over against our concept of God. The real question, therefore, is whether an analysis of the concept of God can provide an adequate basis for asserting God’s existence—for positing an object that falls under our concept of God. Now the critics may mean to argue that no amount of conceptual analysis can ever provide such a basis. If so, however, their “argument” is merely a *non sequitur*. For we may grant their basic point that existence is difference from “real” predicates, and that in predicating existence we are not *merely* making an assertion about our concept of an object. It simply does not follow, however, that we cannot, through the analysis of a concept, find an adequate basis for positing an object that falls under that concept. The critics may well insist on their point. And we may even be inclined—“intuitively”—to agree with them. But so far, this has simply not been shown.

This is not to say, of course, that the first part of the ontological argument (as normally developed) has now been vindicated. Nor is it simply to dismiss the standard line of criticism. This line of criticism seems, at the very least, to raise an important question. For given the difference between existence and other predicates—those predicates which Kant calls “real” or “determining”—we must surely wonder how it is possible, by analyzing a concept, to find in it an adequate basis for positing an object over against the concept. The point is, however, that the standard line of criticism does not suffice to *show* that this is not possible. And if we are to reject the argument, therefore, we shall have to have better grounds than those usually advanced.

Such grounds can, I think, be found. In fact, a suggestion as to what they are can be found in Kant’s own discussion of the ontological argument. This may come as something of a surprise, for as I have already indicated, Kant’s attack on the

<sup>9</sup> *Critique of Pure Reason*, tr. N. K. Smith, second impression (London, 1933; reprinted 1963). A599=B627. (I shall refer to the *Critique* in the usual way, citing the pagination of the A- and B-editions.)

ontological argument is usually taken as a paradigm of the standard line of criticism that we have just been considering. There is, I believe, considerable textual basis for interpreting Kant's criticism along the standard lines. But the text also points in other directions. In at least one place, Kant seems to be trying to develop an objection stronger than the standard one. He argues:

If, in an identical proposition, I reject the predicate while retaining the subject, contradiction results; and I therefore say that the former belongs necessarily to the latter. But if we reject subject and predicate alike, there is no contradiction; for nothing is then left that can be contradicted. To posit a triangle, and yet to reject its three angles, is self-contradictory; but there is no contradiction in rejecting the triangle together with its three angles. The same holds true of the concept of an absolutely necessary being. If its existence is rejected, we reject the thing itself with all its predicates; and no question of contradiction can then arise.<sup>10</sup>

Now there may well be some question about just what Kant means here. He may simply be arguing that we cannot contradict ourselves when we reject the existence of an object, since in affirming or denying the existence of something we are not making any claims about how an object is or is not conceived—rather, we are simply positing or refusing to posit an object for the concept. In that case, of course, his point is simply a variation on the standard line of criticism, and our response to it would be basically the same.

The passage seems to call for a different interpretation, however, since Kant seems to be trying to make a stronger point than this. His argument appears to run along the following lines: To contradict ourselves we must first posit something as a subject of predication. When we deny the existence of God, however, we do not posit a subject of predication. But then we cannot be contradicting ourselves when we deny that God exists, and the judgment that God exists cannot be analytic.

If this is in fact the thrust of Kant's objection, then his argument seems to rest on certain presuppositions concerning judgment in general. Filling these in, his argument appears to take the following form: Whenever one makes a judgment with a real predicate—a judgment of the form, *S* is (or is not) *P*—he posits an object conceived in a certain way and attributes (or denies) a certain

predicate to it. All such judgments, that is to say, involve the positing of an object, the affirmation of the existence of something. Indeed, the point is even stronger than that. The positing is not something distinct from the judging—not something merely presupposed by or implied in it. To judge is to attribute (or deny) a predicate to something, and hence the positing is an integral part of the judging itself. Moreover, this positing is just what we mean by an affirmation of existence. To affirm the existence of something is nothing more (nor less!) than to take that something as a subject of judgment—to take it as something to which we may attribute predicates truly or falsely.

In particular, Kant seems to be arguing that we cannot make sense of an analytic judgment unless we see it as involving the positing of a subject of predication. In judging that *S* is *P*, where *P* is contained in the concept *S*, one posits the subject, according to Kant, and then attributes to it a predicate which has already been thought (implicitly) in our original concept of it. Similarly, in denying that *S* is *P*, where *P* is contained in the concept *S*, one posits the subject, but then denies of it a predicate which has already been thought (implicitly) in our concept of it; and this, of course, is self-contradictory. Kant's point is that if no object were posited, there would be nothing of which we could deny the predicate, and hence there could be no self-contradiction. As he puts it: "If its existence is rejected, we reject the thing itself with all its predicates; and no question of contradiction can then remain."

From these remarks about judgment in general, and about analytic judgments in particular, it follows that the first part of the ontological argument, as it is normally interpreted, rests on a basic mistake. When we deny that God exists, we do not posit the subject (God) and then deny a predicate (existence) to it. Rather, we simply refuse to posit the subject in the first place. Since nothing is posited, however, there is nothing to which a predicate may be attributed or denied. Consequently there can be no question of our contradicting ourselves. And since the judgment that God does *not* exist is not self-contradictory, the judgment that he *does* exist cannot be analytic.

Now one may well wonder whether Kant actually means to argue in the way I have suggested, and whether the general account of judgment that I have just attributed to him is plausible. I am

<sup>10</sup> *Critique of Pure Reason*, A594-5 = B622-3.

inclined to think that the interpretation I have suggested fits the text quite well. I am also inclined to think that the general account of judgment is a reasonable one, for I think it connects the notions of existence and judgment in the proper way. (A judgment, it seems to me, is an attempt to think truly about something; and as such, I think it must involve the positing of an object with respect to which our thinking is true or false. Moreover, it also seems to me that when we posit the existence of something we must be taking that something as a subject of judgment—as something to which predicates may be attributed, truly or falsely; for otherwise, our positing would have no content.) I shall not try to argue these points now, however. For my primary interest here is to show what results from this line of attack on the ontological argument.

Assuming, then, that my interpretation is correct, I think that Kant has shown that the judgment that God exists cannot be analytic, and I think he has done so by appealing to a general and plausible theory of judgment. He has successfully exposed the error which underlies the first part of the ontological argument, as it is usually developed—the error which we suspected was there, but which we found difficult to pin down. I do *not* think, though, that this is the end of the story. I think it is possible to develop the first part of the ontological argument in a way that escapes Kant's criticism. And ironically, I think that Kant himself provides a suggestion as to how this might be done.

Kant has shown that the judgment that God exists cannot be analytic, and he means to conclude that the ontological argument is therefore refuted. But he adds the following remark: "The only way of evading this conclusion is to argue that there are subjects which cannot be removed, and must always remain."<sup>11</sup> Unfortunately, he dismisses this reply out of hand. To claim that there is an irremovable subject is no different, he says, from claiming that there is a being whose existence cannot be denied without self-contradiction; and that, as he has just shown, cannot be so. I believe, however, that Kant is wrong about this. I think it is possible to make sense of the notion of an irremovable subject—a subject which we cannot refuse to posit. And I think we can do so without having to fall back on the admittedly erroneous claim that the judgment that God exists is analytic.

What I mean to suggest is very likely apparent by now. It is that substance, as Spinoza conceives it, is just such an irremovable subject. Hence if we accept Spinoza's metaphysical view, we cannot intelligibly refuse to posit substance—we cannot intelligibly doubt its existence.

The point is actually quite simple. Given Spinoza's theory of substance, substance is, as we have seen, that which is metaphysically fundamental. It is that of which all other things are mere modes, while it itself is not a mode of anything more fundamental. But this is simply to say that substance is, on Spinoza's view, reality; for all things are real by virtue of their being comprehended within substance. To doubt the existence of substance would therefore be to doubt the existence of reality itself. A doubt of that sort would simply be unintelligible, however, for the doubt would make its own resolution impossible in principle. After all, we can resolve doubts about whether something exists only by appealing to other judgments (which we take to be true) about what really exists. If we were to doubt, for example, whether chlorine exists, our doubt would have to be resolved by appealing, e.g., to the judgment that there exists a certain chemical substance, that this substance has an atomic weight greater than that of sulphur but less than that of argon, that this substance exhibits appropriate similarities to other elements in group vii of the periodic table, etc. If we were to doubt whether reality itself exists, however, we would automatically rule out any such appeal to other judgments about what really is. In raising the doubt, therefore, we would make it impossible in principle ever to resolve it. And thus what seemed at first to be an intelligible question turns out, on reflection, to be quite unintelligible.

This same point may also be approached from another direction. If it is true that we cannot intelligibly doubt the existence of substance, then for similar reasons, the assertion of the non-existence of substance is unintelligible. This point can be made with terminology borrowed from Kant, and it will allow us to see more clearly why Spinoza's concept of substance turns out to be the concept of an irremovable subject.

Kant argues that there can be no self-contradiction in simply refusing to posit an object. I think, moreover, that he is correct about this. Hence I agree that the denial of the existence of substance

<sup>11</sup> *Critique of Pure Reason*, A595=B623.



cannot be unintelligible by virtue of being self-contradictory. There is another, and still more fundamental type of absurdity or unintelligibility, however, and we lapse into this type of absurdity when we attempt to deny the existence of substance. The point hinges, of course, on the fact that substance is that which is metaphysically fundamental. In the case of things which are not fundamental, it may well be foolish for us to deny their existence; for in doing so we may contradict the obvious testimony of experience, and we may also contradict a wide range of judgments of whose truth we are quite certain. Still, such denials are not unintelligible. In denying the existence of substance, however, we would not merely contradict the obvious testimony of experience, nor would we merely contradict some limited range of judgments of whose truth we are quite confident. We would rather contradict all judgment whatsoever! For given that substance is the whole of reality, in denying the existence of substance we would be denying the existence of reality itself—i.e., we would be refusing to posit anything whatsoever as a subject of judgment. In doing this, however, we would be rejecting the very possibility of judgment, true or false, for we would be denying that there is anything at all of which our judgments might be true or false. To deny the existence of substance would therefore be absurd or unintelligible in that it would be to reject the very possibility of true or false judgment.

Someone might object that if we deny the existence of substance, there would still be one judgment which is either true or false—namely, the judgment that substance doesn't exist. Hence we would not have rejected all possibility of judgment, nor would we have lapsed into any absurdity. This objection is misguided, I think, but it serves at least to show that the last argument must be formulated a bit more clearly.

The point that I am trying to make will perhaps become clearer if I connect it with some comments on existential judgments. Kant's view of existential judgments contains two basic points, as we have seen. He holds that in making an existential judgment we do not attribute a predicate to an object; rather, we simply posit or refuse to posit an object corresponding to our subject concept.

He also holds that we can always refuse, without absurdity, to posit an object—that there is no such thing as an irremovable subject of judgment. If we assume that Spinoza's theory of substance is correct, however, then it follows that Kant is wrong on both points. For it follows from Spinoza's view that existential judgments do not generally fit the model that Kant sets forth. And it also follows that in the one case which does fit Kant's model, we cannot intelligibly refuse to posit the subject.

If Spinoza's view of substance is correct, then to assert that something exists is ultimately to make an assertion about substance.<sup>12</sup> To assert that water exists, for example, is to assert that substance is so modified as to constitute water. Likewise, to deny that something exists is also to make an assertion—ultimately—about substance. It is to assert that substance is not modified so as to constitute the thing in question. It follows, therefore, that Kant was wrong. Existential judgments, properly understood, do *not* involve a mere positing or refusal to posit. For in either case, whether the existential judgment be affirmative or negative, the judgment ultimately involves the attribution or denial of a predicate to a subject.

On Spinoza's view, then, substance is the ultimate subject of predication, and all judgments must be understood as positing substance, and then attributing or denying predicates to it. All judgments, that is, except those affirming or denying the existence of substance itself. For obviously, neither of these could reasonably be understood as attributing or denying a predicate to substance. They seem rather to fit the model which Kant proposed for existential judgments generally: To affirm the existence of substance is, it seems, simply to posit it while to deny its existence is simply to refuse to posit it. Hence, though Kant's model does not hold good for existential judgments in general, there is one case in which it is applicable. Ironically, however, in this one case, his other main thesis turns out to be false; for it is not possible to refuse to posit substance—not, at least, without lapsing into unintelligibility. To refuse to posit substance would be to refuse to posit the ultimate subject of all predication, and thus to reject all possibility of judgment, true or

<sup>12</sup> I stress the importance of the "ultimately." I am not saying that when people make existential judgments, either in ordinary contexts or in the context of scientific inquiry, they think of themselves as making assertions about substance. Indeed, I shall admit that we can "get along," both in everyday life and in scientific theorizing, without having to employ the concept of substance—i.e., without having to do metaphysics. But when we do adopt the standpoint of the metaphysician, we have to regard all existential judgments as assertions made about substance—about that which is ultimately real.

false. Furthermore, it cannot be objected that the denial of the existence of substance would itself be true or false. For in refusing to posit substance we would not, properly speaking, be making a judgment. We would not be asserting that reality is constituted or arranged in a certain way. Rather, we would simply be rejecting all possibility of making a judgment that is true or false.

Given Spinoza's theory of substance, then, it follows that we cannot intelligibly doubt the existence of substance. Assuming, furthermore, that Spinoza is right in thinking that substance has to be conceived as God, it follows that we cannot intelligibly doubt the existence of God. Putting this conclusion together with that of the previous section, we see that Spinoza's theory of substance does indeed provide a basis for constructing a defensible version of the ontological argument. For his view implies that we cannot intelligibly doubt the existence of God, and that God exists necessarily.

#### IV

There are many objections which might be raised against the argument of the preceding two sections. In this last section I shall try, briefly, to anticipate some of them.

(1) First, let me stress that I do *not* think I have defended the ontological argument against *all* possible objections. I have only tried to show that *if* Spinoza is right in his theory of substance and his assumption concerning causality, and *if* he is also right in thinking that substance must be conceived as God, then it follows that we cannot doubt God's existence, and that he exists necessarily. I have not tried to defend either of these assumptions. I believe, nonetheless, that the point I have tried to make is important; for it serves to refute the most common—and the most radical—sort of objection to the ontological argument. Critics of the argument have typically argued that no matter how God is conceived, we cannot establish any *a priori* link between our concept of God and an object corresponding to that concept. I have shown, I think, that this radical criticism is incorrect.

(2) This suggests a further question. If we can doubt Spinoza's metaphysical view, then can't we doubt the existence of Spinoza's God? And doesn't this serve to show that my defense of the first part of the ontological argument is wrong?

The answer to the first question is "Yes." We *can* doubt the existence of Spinoza's God. We can do so, however, *only* by doubting his metaphysical view. The answer to the second question is therefore "No"; for what I have tried to show is precisely that *if* Spinoza's metaphysical view is correct, then it follows that God is not the sort of being whose existence can be intelligibly doubted.

(3) As I suggested earlier, my defense of the ontological argument has an interesting consequence. What I have done, in analyzing Spinoza's view, is to break it down into two components. First, there is a core argument—what we might call the ontological argument *proper*—which shows that we cannot doubt the existence of substance, and that substance exists necessarily. Secondly, there is a surrounding body of doctrine in which this core is embedded—a body of doctrine which purports to state how we must conceive of substance. (In Spinoza's case, the doctrine is that substance must be conceived as God.) If I am right about this, then it follows that the ontological argument proper does not deal with God, but rather with substance; and it is a further question whether substance has to be conceived as God.

It might be said, therefore, that the ontological argument, as I interpret it, is in a certain sense "general." It shows that regardless of how one conceives of substance—whether as God, as matter, or in some other way—he will be committed to the ontological argument proper so long as he employs the concept of substance. We can set aside questions about the specific way in which substance is to be conceived and simply reason about substance. And when we do, we see that if we accept a metaphysical view which employs the concept of substance, then we must also accept the ontological argument, in both of its parts.

(4) The ontological argument seems generally to be thought of as an attempt to demonstrate that God exists, and indeed, that it is necessarily true that God exists. What I have argued in Sect. III is that given a certain concept of God, it follows that we cannot intelligibly doubt his existence. I have therefore treated the argument as something less venturesome than an attempt to demonstrate that God exists. I think this is quite appropriate, however. I realize that some proponents of the argument (notably Descartes) have regarded it as providing a demonstration of God's existence; and I also realize that this is what the critics of the argument have usually taken it to be. Still, I think it is unfortunate that the argument should

be thought of in this way. And I doubt that either Anselm or Spinoza thought of it in this way.

It is more accurate, I think, to regard the ontological argument as something like a corollary. For what the argument does, as I see it, is simply to spell out the consequences of a certain metaphysical view—the theory of substance formulated by Spinoza. The interesting question, then, is

not whether the ontological argument is correct, but whether the metaphysical view on which it rests is the correct metaphysical view. It is in dealing with this question that the proponent of the ontological argument must undertake his labor of inquiry. And it is only by answering this question that we would be able to assess the full force of the ontological argument.

*University of Kansas*

*Received March 22, 1973*

## IV. PETITIONING GOD

ROBERT YOUNG

IN the Judaeo-Christian tradition prayer to the Godhead ranges over petition, entreaty, ex-postulation, confession, thanksgiving, praise, adoration, meditation, and intercession. Undoubtedly, most of the difficult philosophical issues which arise in connection with this practice of prayer do so in connection with petitioning God, and especially when the petitions are for what are sometimes called "particular providences," rather than for forgiveness for the believer. (The discussion in Sect. II below will show, I think, that this is not always a suitable label.) In such circumstances the practice raises issues about the concept of "miracle," about God's providential arranging of the world, and about the effects, if any, which human freedom has on such arrangements. I propose to consider difficulties of these kinds associated with the practice of petitionary prayer.<sup>1</sup>

Before proceeding to do this, I should point out that I think it is largely true that what one prays for and how one conceives the whole activity of prayer reflects accurately one's conception of God and hence of that personal relationship with God which the activity presupposes. In light of these remarks I shall not be concerned with views of petitionary prayer which seek to collapse it into some other part of the range covered by the blanket term "prayer." Views of this latter form replace the concept of God found in the Judaeo-Christian tradition with an empty shell and are not in my view sufficiently interesting to warrant further discussion here.

### I

Though this places God under no constraint to answer positively, it is normative in the tradition with which I am concerned, that prayer be in accord with the divine will if it is to induce the desired response from the Godhead. (It is God's graciousness and faithfulness to his promises, not the petitioner's rights, that underpins answers to prayer.) When a request does not accord with

God's will it is this lack of accord which is said to explain why such a petition received a negative response. One could cite here such instances as Moses' petition to enter Canaan or Paul's request for the removal of his "thorn in the flesh." Similarly, when it is said that a prayer of Daniel's was answered before he could even phrase it, this is to be explained presumably by the fact of God's knowing that Daniel's intentions were substantively in accord with the divine will.

Saying that prayer must accord with the divine will does not seemingly commit one to saying that this will must in turn be accommodated solely to non-miraculous occurrences. Hence, petitionary prayers may consist in prayers for the happening of a miracle. That this is envisaged in the tradition should be plain from such instances as Elijah's prayer for God to declare his living power by igniting the wet offering on Mt. Carmel and Peter's prayer for healing the man lame from birth whom he encountered in the Temple. Hence there would seem to be no grounds consistent with these narratives, other than doubts about the content of the divine will, for holding that because a prayer for a certain event would apparently be a prayer for a miracle, one should either refrain from praying for the event or not believe God would answer such a prayer positively. (I presume, of course, that any other requisite conditions such as the one praying being a man of faith, having a fundamentally sound attitude to the nature of the privilege of prayer, and so on are all fulfilled.)

Clearly then, it is desirable that a believer in the efficacy of petitionary prayer concerning the miraculous come up with an account of miracle worthy of the name, for the concepts of the miraculous and of petitionary prayer will at times be inseparably linked.

It should be a matter of regret to the Judaeo-Christian theist that the concepts of the miraculous one finds bruited about are not altogether satisfactory. I shall briefly show that two commonly

<sup>1</sup> Any reader interested in such views could consult two of the best known accounts, namely T. R. Miles, *Religion and the Scientific Outlook* (London, 1959), ch. 12; and D. Z. Phillips, *The Concept of Prayer* (London, 1965), ch. 6.

advanced approaches are less than successful and then put forward what I trust is an improved account. I shall then discuss the relation between the concept of "miracle" advocated and petitionary prayer when such prayer involves a request for a miracle.

Consider first of all what has been called the idea of a "coincidence" miracle.<sup>2</sup> Holland illustrates the notion by devising a case in which a child strays onto a railway line unaware that a train is approaching around a curve where there is no chance of the driver seeing the child in time to stop. The mother, watching from a distance and unable to help, sees the train approach and grind to a halt a few feet from the child. Holland writes:

The mother thanks God for the miracle; which she never ceases to think of as such although, as she in due course learns, there was nothing supernatural about the manner in which the brakes of the train came to be applied. The driver had fainted, for a reason that had nothing to do with the presence of the child on the line, and the brakes were applied automatically as his hand ceased to exert pressure on the control lever. (P. 43.)

There are three points I wish to make about such a view of the miraculous. It is important to see, first, that, on this view, the factors bringing about the event do not include specific initiatives on God's part, though his general agency, which is a necessary condition for the occurrence of anything at all, is present. Any distinction between what theologians call God's *general* providence and his *special* providence lapses on this view.

Secondly, any coincidence theorist is required to offer some epistemic support for the claim that only certain (non-disastrous) coincidences constitute miracles. Presumably, a coincidence theorist who really believes in a transcendent God would make an appeal to the goodness of his God. This reply has some force but not a great deal when one keeps to the forefront of one's thinking the contention that such a God hasn't done anything to bring the "miracle" about, apart from what-

ever he may have done in bringing into existence the whole show with whatever purpose it has built into it.

Thirdly, and most significantly, this conception of the miraculous cannot do justice to the sorts of events exemplified in narratives of miracles contained in the Judaeo-Christian tradition. It would be unfair to Holland not to point out that he recognizes this.

I turn, therefore, to the dominant view of miracle—that which involves a violation of the laws of nature by God. The first thing to notice is that the idea of a violation of a natural law is parasitic on the conception of natural law. On the face of it, because natural laws are descriptive not prescriptive, the concept of a natural law is to be filled out in terms of what (and only what) things happen or are done. This being so, any event which appears to be different in character from what has hitherto been observed cannot consistently fit the two descriptions: "event which occurred at *P* at *t*" and "event which violates (a stated) law of nature."

Advocates of this conception of the miraculous have tried to get around this difficulty. Indeed it is developments of the violation concept in response to this sort of problem which have enlivened recent discussions of the concept of "miracle."<sup>3</sup> Smart introduced the claim that a miraculous violation of a law of nature is an occurrence of a non-repeatable counter-instance to such a law. Swinburne focuses on a question which draws out some of the implications of this claim. He writes:

But what are we to say if we have good reason to believe that an event *E* has occurred contrary to predictions of a formula *L* which otherwise we have good reason to believe to be a law of nature and we have good reason to believe that events similar to *E* would not occur in circumstances as similar as we like in any respect to those of the occurrence of *E*? (P. 27.)

Swinburne points out that we could *either* say that *L* cannot be the law of nature operative in the field *or* that *L* is the operative law but that an exceptional non-repeatable counter-instance to its occurrence has occurred. He, like Wallace and

<sup>2</sup> For a contemporary statement see R. F. Holland, "The Miraculous," *American Philosophical Quarterly*, vol. 2 (1965), pp. 43-51 (esp. pp. 43ff.). Cf. also Mary Hesse, "Miracles and the Laws of Nature," in C. F. D. Moule (ed.), *Miracles* (London, 1965).

<sup>3</sup> I have in mind Ninian Smart, *Philosophers and Religious Truth* (London, 1964), ch. 2; Holland, *op. cit.*, p. 49; R. G. Swinburne, *The Concept of Miracle* (London, 1970); R. C. Wallace, "Hume, Flew and the Miraculous," *The Philosophical Quarterly*, vol. 20 (1970), pp. 230-243.

Tan Tei Wei's "Recent Discussions of Miracles," *Sophia*, vol. 11 (1972), pp. 21-28, seeks to go beyond these other accounts while remaining attached to this general approach.

Smart and like Tan Tei Wei when the event has occurred in connection with an ideological personality and relevantly supports his ideology, believes we should jump the latter way. Now proponents of this view hold that jumping the way they advocate requires holding the occurrence to be scientifically unexplainable. For it is contended that it makes sense to suppose that on occasion the physically impossible occurs, or, to put it as Holland does, one must embrace simultaneously the empirical certainty and the conceptual impossibility of a violation of a law of nature.

It seems, first of all, that this view of a miracle leads in the direction of fideism. The view, in other words, seems to hinge on taking the strength of one's rational belief in the occurrence of the event *E* to be such that one can hold the further belief that any modification which was made to the laws of nature to enable them to predict *E* might not yield any more successful predictions than *L* (currently thought of as the pertinent law of nature). Furthermore, one needs to hold that any modification might make the new law(s) so clumsy that there would be no reason to believe their as yet untested predictions would be successful. Clearly a good deal is being made to hang on conjecture about the availability or otherwise of a simple, elegant, and yet more comprehensive law than *L*. It is important that I point out that I do not contend that proponents of this view must beg the question by asserting the incorrigibility of all claims about what does or does not violate the laws of nature. Swinburne is clear that our reasons for believing *E* happened may be good without being logically conclusive. Now in all probability the response of an unbeliever to all this will be to counter with the suggestion that a number of rationally defensible accounts will be describable that show it would be more natural to believe a better law is available than that the existing one was adequate but yet had been unrepeatably violated. Such rationally defensible accounts could appeal to factors like the paucity of our imagination in not as yet devising a better law or the smallness of the number of instances relevant to establishing *L* and so on. There is an obvious parallel with the reaction to be expected of a macro-level determinist confronted with an apparent instance of macro-indeterminacy.

The unbeliever has an even stronger bit of ammunition at his disposal in this connection, though. He might wish to reject the account of a law of nature involved in the concept of a miracu-

lous violation. This for the reason that genuine nomologicals are normally distinguished from mere *de facto* universal generalizations by their capacity to support counter-factuals. Thus if *E* was judged to have occurred, it would be contended, I should think, that *L* was a mere accidental universal and thus that it could not be the relevant law of nature. Fideism of the kind I think is involved here does not seem to be a desirable theological niche.

I turn to a second criticism. Holland holds that on pain of foregoing a proper understanding of miracles as violations, one must accept the *paradoxical* consequence that on occasions the physically impossible occurs. Swinburne does allow, though, that those who find such an idea hard to take might prefer a revised notion of "physical impossibility." Unfortunately, this move founders on the fact that we are quite happy with the claim that something is physically impossible *simpliciter* if it is incompatible with the laws of nature. (I return to the need for the italicized rider below.) In order that one feel any degree of constraint about revising the notion of the physically impossible, a good reason is surely required. The only candidate in Swinburne's discussion is his contention that the concept of a violation of a law of nature is coherent, that is, logically possible of instantiation. Indeed, it is urged that if it were not, no ready sense could be given to the notion of a violation of a law of nature. One might retort that it is precisely because no ready sense can be given to the notion of a violation of a law of nature that it is an incoherent concept.

It seems likely at this stage that debate will break down, with one side claiming the coherence of the concept and the other denying it. Interminable swapping of intuitions about whether the notion of the physically impossible can be stretched in the way Swinburne envisages seems likely to be a sterile business. I shall try to escape sterility by outlining a different approach altogether. I am heartened in this undertaking by the evidence from biblical theology that there are no entailments about the overriding or violating of a natural order in reports of miracles contained in the documents of the Judaeo-Christian tradition.

Consider what is likely to happen if the predicted value under a certain law does not match the actual value. Allowing that observations are reliable, any of the following five alternatives seems open:

- (i) a search for discrepant factors (e.g., among the tacit conditions) may be instigated;

- (ii) a search for a mistake in the specification of the independently specifiable initial conditions may be instigated;
- (iii) a search for the features of the theory, say, being tested which would be more readily revisable than firmly entrenched high order generalizations associated with it, might be instigated;
- (iv) a revision of the law (or theory based on the law) might be undertaken by so expanding it as to accommodate the contrary instance within the revised version;
- (v) rejection of the law may become unavoidable.

Because very many supposed laws of nature provide only vague specifications of, for example, the initial conditions, criteria of disconfirmability for such laws are also vague. It should come as no surprise, therefore, when dealing with laws in which the antecedent does not specify a genuine sufficient condition of the effect (the other conditions being only tacitly specified), that given an instance of the stated antecedent the absence of the consequent does not directly disconfirm the law. The possibility remains open that one or the other of the tacit conditions remains unsatisfied. Thus if one were to holiday in a mountain resort one should remain undismayed upon checking the temperature at which the water boils to find that it does not do so at 100° C. Once reminded of the presence of the new factor in the water-boiling situation—height above sea-level—one recognizes that the relevant law about the boiling point of water has not been violated just because of the change in the initial or surrounding conditions.

It is my contention that the most satisfactory account of the concept of "miracle" is obtained by considering God to be a (specifically rather than just generally) active agent-factor in the occurrence of a miraculous event such that his doing something he does not normally do introduces a new (and possibly unique) set of causally sufficient conditions. Here there is no violation of law, nor is there a religiously interpreted coincidence. More fully: when a miracle occurs, God is active as a factor in the causally operative set of factors (out of the perhaps several possible sets sufficient for the event's occurrence) and his activity alters the outcome from what it perhaps

would have been if, contrary to fact, he had not been actively involved. It is open to the theist on this account to say, furthermore, that had God not actively involved himself, the (miraculous) effect would not have come about. I should acknowledge that this account requires my rejecting the view that God is a specifically (not just generally) active agency in each and every event. I think such a view is incompatible with central features of Judaeo-Christian theism like God's steadfast disposition to do the right. Likewise, the account is implacably opposed to deism and popular modern variations on the theme.<sup>4</sup>

I suggest the following sketch, therefore, as a description of what the believer may be doing when he petitions God for something which seems very unlikely to happen. What he may be doing is asking that God respond positively to his petition by actively engaging himself in the task of bringing about something that perhaps would not otherwise occur. Suppose, for example, that a believer prays for the removal of an inoperable and otherwise medically intractable tumor. He is on my account—and I think this account concurs with orthodox belief—almost certainly asking that God alter the surrounding conditions of the state of affairs by "making his presence felt." It is important that the reader notices I have said that the believer is "almost certainly" doing this and that what he asks for may "perhaps not otherwise occur." These qualifications are not intended to take the bite out of my earlier remarks but are intended to point to certain important theological considerations which cry out for attention at this juncture.

These considerations arise on two fronts. First, those of an epistemic kind and, secondly, those about God's providential arranging of events. It will facilitate progress if these two categories are further divided. The epistemic considerations may be split up between those that might arise for believers and those that divide believers from unbelievers. The considerations about providential arrangements require keeping distinct the sort of problem raised if the world has been predetermined in its causes by God, from the sort to which a non-predetermined world would give rise.

On the first count there are basically three claims that compete for our assent and which we must try to assess. It is to be taken as true *ex hypothesi*

<sup>4</sup> I have tried to say a little more about problems raised by the account in "Miracles and Epistemology," *Religious Studies*, vol. 8 (1972), pp. 115-126.

that the believer prays for a particular event, *E*, to come about regardless of which claim is held in the end rationally to be the strongest. First, there is the claim that *E* was miraculously brought about by God in answer to prayer. Next, there is the claim that *E* was brought about non-miraculously by God, that is, God contrived to bring it about other than by acting as a new causally significant factor in the surrounding conditions. Thirdly, there is the unbeliever's claim that God could not have brought about *E* either miraculously or non-miraculously for the reason that he doesn't exist. A certain artificiality in keeping separate the epistemic and the providential considerations is, no doubt, evident, because the first two claims are obviously tied up with God's providential activity. But for ease of exposition I shall try to ignore this artificiality as much as is possible.

The response made by a believer in supporting one of the former two claims will depend very much on the nature of what is prayed for. Suppose the believer prays, as in my earlier example, for the removal of an inoperable and otherwise medically intractable tumor, which the medical profession is no longer treating and suppose, further, that subsequent to the time of praying medical reports indicate that the malignant tumor has vanished. Suppose, further, that no reasonably plausible explanation is to hand which involves reference only to unusual but not entirely dismissable this-worldly curative factors like, for example, the herbal diet of the subject or psychosomatic healing processes. Faced with such a situation a believer may well contend that not only is it entirely reasonable to conclude that God's miraculous activity is the most plausible explanation available, he may hold further that for either other believers or unbelievers not to adopt such an explanation would be quite unreasonable.

By contrast consider the following case which has been suggested to me. Imagine an astronomically ill-informed missionary who prays in the presence of the heathen for God to show a sign in the heavens in order that the heathen may see God's glory, turn to God and be converted. Even as he prays the sun is eclipsed. The very time of the eclipse is that predicted by astronomers. The missionary may contend that the eclipse was a miraculous answer to prayer occasioned by God's

active causal agency in the (altered) surrounding conditions. Other believers may, however, contend that our knowledge of astronomy shows that there is discoverable a complete explanation of the heavenly event *and* its timing,<sup>5</sup> that such an explanation involves reference only to non-redundant causal factors and that any reference to God's miraculous activity is ruled out as implausible and unnecessary. Adoption of such a view would in no way prejudice such a believer's commitment to theism. It would necessitate his refraining from claiming, though, that the eclipse had any *special* revelatory rôle. This points up an important difference as between the two cases of answered petitionary prayer outlined, namely, that if in the case of the healing it is held that the most plausible explanatory hypothesis involves reference to God's miraculous doings, it may further be held that the healing does have special revelatory significance. It is important to bear in mind, though, that viewing the miraculous in this way is to view it as partially constituting the total revelation rather than as proving or endorsing that revelation.

Clearly the unbeliever is able to argue more convincingly in the second than in the first case, that talk of God positively responding to the petition offends against a properly parsimonious attitude. At the same time it is clear that he must also try to collapse the sort of explanation offered in the first case into that of the second on pain of unreasonably holding out against a (more plausible) theistic explanation. This follows from his aiming to show that in neither case did the course of events turn out as the petitioner requested *because* (in part) he requested it. I shall return to this shortly.

Before doing that I should point out that believers sometimes neglect a not too dissimilar difficulty. The difficulty is illustrated by cases in which people pray to what the Judaeo-Christian tradition speaks of as "no-gods" and yet appear to receive positive responses requiring explanations approximating to the idea of the miraculous I have employed. Faced with such cases, the believer seems committed to doing one of two things. He may, first, try to account for the "answer" in terms of the activity of his own Judaeo-Christian God while contending that the petitioner has an inadequate theological recognition

<sup>5</sup> It is crucial that the timing be mentioned. Believers sometimes respond to the suggestion that there are completely adequate this-worldly causal explanations with the counter that consideration of the timing is what reveals the event in question as a genuine answer. The rider I have added defuses such a counter.



of that God. Or he may, secondly, contend that there is available some more plausible explanatory hypothesis which makes reference only to this-worldly causal factors (e.g., chicanery or the overlooking of some relevant variable). What he probably is precluded from arguing is that the occurrence shows how much more "mysterious" the world is than we often allow. Because if he says this kind of thing he lays himself open to an *ad hominem* against his own endeavors to account for what he believes are events requiring explanatory reference to God's agent-causative activity.

It is time to return to the point made above about the unbeliever aiming to show that events do not come about because (in part) believers petition God. In dealing with the difficulties associated with God's providential arranging of the world tied up with this claim, I shall draw on the previously outlined division of possible worlds, which a theist needs to consider. Namely, ones predetermined in their causes by God and ones not so predetermined, that is, ones which involve some element of indeterminacy.

## II

Most philosophical critics of Judaeo-Christian religious beliefs, as well as most defenders, deny the consistency of such beliefs with the idea that God has predetermined the world in its causes. They do this chiefly because they hold that such a world would preclude the freedom of action and decision required by other features of the Judaeo-Christian vision of the world (e.g., the alleged culpability of human sinfulness). Nevertheless, there have always been theological determinists and any attempt to outline a defense of Judaeo-Christian petitionary prayer must pay some attention to their views. (It so happens that I think many dismissals of theological determinism are altogether too swift and that theological indeterminists don't always clearly perceive the difficulties they inherit from the doctrine of indeterminism itself. But this is itself a complicated issue which can't be pursued here.)

I think three kinds of circumstance need to be investigated in any consideration of petitionary prayer in a world God has predetermined in its causes. First, there is the circumstance in which the event prayed for is entirely explicable in terms of non-redundant this-worldly causes. The believer who holds that God *has* answered his prayer in such a circumstance must, I take it, be asserting

that there is over-determination. Thus he holds that the event would have come about even if the (causally determined) prayer had not been offered, *but* that had God not predetermined the natural causes of this event, God would himself have acted as a causal agent to bring about a miracle. Such a bold claim would require a high degree of certitude about God's will in the matter. Whether and how the believer could come into possession of such certitude is another matter—its treatment would require too great a digression justifiably to be included here. Any believer who did not stick to the claim that God had answered such a prayer would not need to run the over-determination thesis, nor would he be precluded thereby from praising God's graciousness.

The second circumstance that can be envisaged is as follows. A believer might hold that God has so predetermined the world that on at least some occasions petitioning God is itself a causally determined factor which, when taken in conjunction with the state of the world at the particular time, constitutes a member of the jointly sufficient set of factors productive of the event.

This contention must not be confused with the quite different one that might be proffered in a case in which one can plausibly connect the praying with the "answering" event as natural cause and effect. This latter circumstance is no doubt one which is more apt to arise from public occasions of prayer, but could also do so on easily conceivable occasions of private prayer. Private prayer on occasions could, for instance, itself effect the answer sought, namely some therapeutic effect on the one praying. Depression, for example, might be thus lifted.

To return to the second of the circumstances isolated. The intricacy of God's provision presumably can extend to what the second circumstance envisages. Critics of theological determinism, of course, respond to this kind of thinking with the claim that it is incompatible with the core Judaeo-Christian concept of human freedom. Once again this brings us up against an issue that cannot be resolved here (and which I am doubtful I can resolve anyway). I can only say that the theological determinist must strive to show that God's predetermining the world does not entail that men are thereby coerced and that such predetermining is compatible with freedom. The case constitutes a special instance of the freedom and determinism controversy.

The third circumstance I wish to discuss is that

in which God brings about a miracle (in part) because of the prayer of the believing petitioner. What this amounts to in a world predetermined in its causes by God, is that the (causally determined) prayer is in God's providence part of the very reason(s) that he has for acting as a causally effective agent in altering the boundary conditions of the world by his presence, and doing something he does not normally do. Again it may be objected that it is not freedom but manipulation that is present here, and again this objection must be resisted by the theological determinist. The theological determinist will contend that the situation just outlined does not undermine direct personal relations with God. Rather, it might be said, God's care in arranging matters this way testifies to the importance petitionary prayer has. What is more, the theological determinist can point to the promise (in a Christian setting, at least) that the Father "knows what (we) need before (we) ask him" (*Matthew 6*) in support of his position.

I turn now to consider a world not predetermined in its entirety by God. One uncompromising defender, Peter Geach, holds, in fact, that no petitionary prayer can be answered by the occurrence of an event that would have happened even if the prayer had not been offered.<sup>6</sup> In Geach's terms, at the time of the prayer the event prayed for must have two-way contingency, that is, it must be possible both for the event to come about and also not to come about. What this entails on Geach's account is that a future event cannot be contingent if, miracles apart, it is already determined in its causes (p. 94). Quite obviously Geach is committed by a claim of this ilk to arguing that the sphere of genuine contingency is very great. Thus if one is to pray earnestly about the weather one must believe that the weather does constitute a future contingent state of affairs. (He appears to believe that while sunsets and eclipses are not thus contingent, that rainfall is.) Geach does not deny there is a great deal of determinism in the material world (p. 97). He does think, though, that it is false that in the realm of natural causes events are *in principle* predictable. He writes:

There are many large-scale future events in the physical world which are contingent in regard to all created causes, and which we cannot predict (even leaving aside the possibility of miracles) from con-

sideration of any created causes; there has to be this element of chance in things if human choices are to have any *Spielraum*, as they manifestly have. But such contingent events do not fall outside the order of Providence, which can arrange them so as to answer prayers. Whether and when prayers are so answered, it is not for philosophy to say. (P. 97.)

I shall not comment on the sentiment expressed in the final quoted sentence, but there are two things about the whole drift of the passage and of Geach's argument that warrant discussion.

First, Geach commits himself to a strong thesis about predictability and its relation to determination. While he may prove to be right about the fundamental contingency of many events in our world he certainly seems to rest his position pretty much on our present *ignorance* about the causes (if any) of some events. But this sort of epistemic consideration does not tell us much about what is ontologically the case. More significantly, for present purposes, his contention that one "can sensibly pray only concerning future contingent issues, where things still can go either way" (p. 93) does not command automatic assent because, in light of the previous discussion of theological determinism, it does seem sensible to pray for events not conceived as Aristotelian-type future contingencies.

Geach's remarks raise a second interesting point, whose treatment must perforce be severely restricted. In the highly developed and powerful version of "the free will defense" offered by Alvin Plantinga in *God and Other Minds* (Ithaca, N.Y., 1967) the libertarian freedom men possess introduces an element of independence *vis-à-vis* God, which places some matters outside God's determination. In the passage cited Geach talks of such matters as free human choices remaining the subject of providential *arrangement*. Consider a situation in which a believer asks God to bring it about that another human act in a particular way (e.g., offer himself for theological training). If the second acts freely in thus putting forward his candidature, this would, according to Plantinga, be something God could not have determined. For Geach, though, it is something God could have arranged. Is it coherent to say that what God can't bring about (namely free human actions) can yet be subject to his arranging? I am

<sup>6</sup> "Praying for Things to Happen" in P. T. Geach, *God and the Soul* (London, 1969), esp. pp. 88ff. His discussion deliberately excludes prayers for miracles and for grace. Terence Penelhum, *Religion and Rationality* (New York, 1971) makes useful comments about aspects of Geach's views not treated here. See pp. 283ff.

reliably informed that he holds the following view which may show that it is. It is Geach's view that God cannot have his purposes thwarted, be surprised or driven to improvisation because he knows all the possible futures. He offers the analogy of a perfect chess player who would know all alternative lines of play and who could, when playing a greatly inferior opponent, announce (prior to the game, I take it) not only that he will win but the particular move by which he will achieve victory. In the chess game there is no doubt that many of the inferior opponent's moves would be forced. He could never be allowed to get into a position where he had options freely open to him, the adoption of any of which would threaten the ultimate manner of victory. To return to the prayer example, there seem to be three alternatives if we allow for the sake of argument that the prayer is said to be answered. First, God might not know the actual outcome until it is realized. Despite distinguished adherents like Boethius and Prior such a view is a denial of orthodoxy. Secondly, God might foreknow that the particular individual's free response *will* fit his salvation purposes (and it is worth stressing that the effects of one individual's training and career may play a far reaching rôle in that scheme). Finally, God might force the individual's hand by overruling his capacity for free choice. The first view doesn't fit the chess situation very well and, furthermore, seems to admit the possibility of God being forced to improvise and to undercut talk of his having arranged the outcome. It is surely more likely that Geach's view is some disjunction of the other two alternatives. The only dispute I would have with him if my preceding speculations are sound is over the propriety of talk of *arrangement* in the event of the second alternative coming about. Without being privy to any more of Geach's thinking there is little more I can say.

I pass on, therefore, to consider difficulties which attach to a recent discussion where a libertarian account of human freedom is adopted, along with a particular view of God's foreknowledge.<sup>7</sup> Penelhum holds that:

... in giving men freedom of choice, God makes it genuinely *uncertain* what they will do, and in consequence (since men's actions are uncertain before they happen) even he does not know what their actions will be before they do them. Yet the fact

that Jones' doing *A* rather than *B* is never *certain* before it happens does not show that his doing *A* is *no more likely* than his doing *B* before it happens. If this is true, there is nothing absurd about the suggestion that the laws of nature incorporate answers to likely prayers. And since it seems necessarily true that more likely things happen than unlikely ones, we have a good reason for expecting that a majority of likely acceptable prayers will in fact be offered and can thus be provided for. (P. 291, Penelhum's italics.)

Before proceeding to the argument I shall offer against this position, I should acknowledge that the talk in the latter part of the quoted passage about some events being more likely may go a little way toward getting an adherent of this position off the hook. Nevertheless, I do not believe it does enough to allay the doubt to which this view is subject. My doubt is simply about the capacity of an adherent of this position to explain how God's uncertainty about what men will do before they act and decide can be consistent with biblical contentions to the effect that certain crucial events in salvation history not only do not come *ex post facto* to God's awareness but are actually ordained by him. Events like the sacking of Jerusalem in the sixth century B.C. and the subsequent exile of the Jews come about, one would think, because of men's free choices. But the prophets and historians of Judah leave the reader in no doubt that none of this complex network of events has gone on without Yahweh's cognizance. I do admit that the Old Testament writers tend to ascribe pretty well all ultimate responsibility for events to Yahweh and hence are fairly uncompromising about the extent of his foreknowledge. But this admission doesn't remove the force of my objection—if anything it strengthens it. The situation in the New Testament corpus is even clearer, I think. Surely the events leading up to Jesus being put to death were not such as to be outside the pale of God's foreknowledge. Thus, for example, Peter's speech (*Acts 2*) draws the two threads of man's action and God's foreknowledge together in a most definite way. Now someone adopting a Penelhum-type stance could retort *either* that none of the crucial events of salvation history stem from the exercise of human freedom and hence present no problem for the analysis of foreknowledge given by Penelhum, *or* that each of the crucial events was so likely relative to other matters that God was able

<sup>7</sup> Cf. Penelhum, *op. cit.*, p. 291. While not a believer, Penelhum thinks his defense of this sort of position is acceptable to both believer and sceptic. I think instances of believers supporting such views could probably be found.

to anticipate their occurrence and provide any needed props. It does not seem to me that either of these possible retorts is reconcilable with mainstream Judaeo-Christian beliefs about God's nature. Eradicating a bulge in one place seems likely to lead to the bulge reappearing elsewhere. The moral to be drawn is, I think, that the Judaeo-Christian theist who believes the world is not entirely predetermined in its causes must try to preserve a view of God's foreknowledge consistent with his actively working his purpose out in the world. This requires, of course, that he also show that such a view of God's foreknowledge is consistent with human freedom as he conceives it. Recent good work on this topic fortunately is encouraging to those who would hope to show the consistency of these doctrines.<sup>8</sup>

*La Trobe University*

Holding these doctrines in tension is vital in the theological indeterminist's program as regards petitionary prayer. Because it would be consistent to hold that God's foreknowledge of men's free decisions to pray, and of their requests, enables him to make intricate provision for the occurrence of the very events that they in fact freely pray for. This will hold whether the events be simply the product of the operation of the natural laws he has established or of his divine agency as an active causal factor in the total set of factors actually sufficient for the occurrence of some miraculous event.

In this essay I have sought to show that petitionary prayer can be understood in a non-reductivist way and remain a coherent activity. The reader must judge how successful the effort has been.<sup>9</sup>

*Received May 14, 1973*

<sup>8</sup> See especially Nelson Pike, *God and Timelessness* (London, 1970), ch. 4 (and bibliographical notes thereto); John Turk Saunders, "Of God and Freedom," *The Philosophical Review*, vol. 75 (1966), pp. 219-225 and "The Temptations of 'Powerlessness'," *American Philosophical Quarterly*, vol. 5 (1968), pp. 100-108. It is of some interest that Penelhum rejects the kind of approach found in the work of Pike and Saunders. Cf. *op. cit.*, p. 290 and ch. 21.

<sup>9</sup> I am grateful to the referee for pointing out some of the errors in a previous version.

## V. APRUDENTIALISM

JOYCE TREBILCOT

Hardly anyone knows how much is gained by  
ignoring the future.

—BERNARD DE FONTENELLE

THIS paper is about living for the moment. This idea has an important place in popular culture, as well as in the social sciences and religion. In sociology, for example, a common theme in discussions of low-income persons is that they have a strong "present-time orientation," due perhaps to inability or unwillingness to control impulses, to defer gratification, and to plan for the future. Indeed, one sociologist has provided an excellent summary of the position to be discussed here. Edward C. Banfield writes of the person who

despite being psychologically capable of providing for the future and despite being in a situation which (as he perceives it) may afford excellent opportunity for investing, nevertheless lives from moment to moment simply because he *prefers* that style of life.<sup>1</sup>

In religion the idea of living for the moment appears, for example, in the Biblical passage which urges one to "take no thought for the morrow; for the morrow shall take thought for the things of itself" (Matt. 6:25-34). On one interpretation this passage means that if one attends exclusively to the present God will provide for the future, so that in living for the moment one is in fact investing; but it is also open to an interpretation which corresponds to the position to be discussed here, viz., that the value of the present is always worth the sacrifice of the future.

Despite the importance of the idea of living in the present, there is little positive discussion of this notion in recent philosophical literature. Aristippus the Cyrenaic apparently held that the goal of action is the pleasure of the moment. But among contemporary ethicists, the most significant contributions to the topic are arguments offered by Thomas Nagel and John Rawls which claim that to prefer one time over another simply

because they are different times is irrational.<sup>2</sup> I shall show that these arguments are unsound.

In this paper then I develop the notion that one ought to live only for the moment as a theory of conduct. In Sect. I, I discuss theories of conduct in general and compare the theory of living in the present with utilitarianism, ethical egoism, and prudence. Sect. II is a detailed explication of the theory itself. Sect. III will answer the charge that "pure time preference" is irrational.

### I

We turn first then to some brief general comments about theories of conduct. The data on which a theory of conduct is based are actual arguments people use about what ought to be done. Such an argument includes a prescription—e.g., "George ought to go to the party"—together with reasons for the prescription—e.g., "because even though there is an inordinately high price to be paid tomorrow for partying tonight, he ought to do what he most wants to do right now." The data on which a theory of conduct is based then are normative arguments. A theory based on such arguments may be used either descriptively, as an explication and summary of the arguments, or normatively, to promote a certain way of life.

Many normative arguments about conduct do not precisely correspond to any theory. When this is so an argument and the prescription which is its conclusion may be given the name of the theory to which the argument most closely corresponds. For example, utilitarianism requires one to consider the consequences of an act for all those affected by it. But a prescription based on consideration of only some of those affected (e.g.,

<sup>1</sup> Edward C. Banfield, *The Unheavenly City* (Boston, 1968), p. 217.

<sup>2</sup> Thomas Nagel, *The Possibility of Altruism* (Oxford, 1970), especially ch. VIII, and John Rawls, *A Theory of Justice* (Cambridge, 1971), pp. 293-295, 420, 423.

only those in one's own group) may be labeled "utilitarian" on the ground that it corresponds more closely to utilitarianism than to any other available theory. Similarly, there are many data for the theory of living only for the moment that are not exact instantiations of it but are greater or lesser approximations to the theory.

Just as normative arguments are labeled "utilitarian," "egoistic," "Kantian," etc., depending on the theory they most closely approximate, so persons are given such labels depending on the theory which most of their arguments approximate. For example, a particular person may usually judge in accordance with utilitarianism, but on some occasions decide that the interests of others ought to be discounted in favor of the agent's own, or that the interests of different people should be given different weights, etc. A person who judges in this way may nevertheless be called a utilitarian on the ground that the person's judgments are usually closer to utilitarianism than to any other familiar theory. It is, of course, possible to base a theory only on the normative arguments employed by a single person over time; indeed, the project of discovering a theory implied by one's own judgments may be a profitable one. But for the task undertaken here, we look to arguments expressing a certain point of view wherever they occur. Hence, if there is no one who always prefers to live only for the moment, this is no objection to the theory. Some of us sometimes do prefer the present, and this is all the data we need.

The arguments which constitute data for a theory of conduct are restricted to those which accurately express the values of the persons who use them. This principle suggests several grounds for the exclusion of arguments from the body of data. First, arguments made on the basis of incomplete or false beliefs about the facts are excluded if there is reason to believe that these arguments would be changed if the subject had more accurate or more complete knowledge. The device of looking to arguments about hypothetical cases, where all relevant consequences of alternative acts are presumed to be known, is useful as a means for guaranteeing that judgments are based on adequate information. Secondly, an argument made when a subject is temporarily in an "abnormal" psychological state—e.g., overcome by emotion, or excessively fatigued—is excluded if there is reason to believe that the subject would repudiate the argument soon after-

wards on the ground that because of the emotion, etc., the argument does not represent the subject's true values. Notice that the criterion for normality here is not what is normal for a particular group but what is usual for a particular subject at that period in life when the argument is used. A third restriction is that arguments made under external pressure—e.g., as a result of threat—are discounted as not representative of the views of the person who makes them. Since our data include only verbal behavior and not other kinds of conduct, we do not need to exclude "compulsive" behavior. If one cannot, because of inner psychological pressures, act in accordance with one's own judgments, we have in any case the data, viz., the arguments. Of course arguments are often themselves influenced by unconscious or unchosen psychological factors; but we are concerned with arguments which accurately express a person's values, not with the origins of those values.

Having made these brief comments about theories of conduct in general, let us turn now to a comparison of the theory of living for the moment with utilitarianism and ethical egoism. Notice first that both utilitarianism and egoism incorporate the principle of prudence. This principle is that one ought to maximize the sum of intrinsic value or minimize the sum of intrinsic disvalue for oneself over one's lifetime as a whole. (I assume here that there is no life after death.) In utilitarianism, prudence is limited by the requirement that one is to consider the well-being of others as well as one's own; one is to promote one's own cumulative well-being only if one has no alternative which would be of greater benefit to the community as a whole. In egoism there is no such restriction, so enlightened egoism is equivalent to amoral or non-altruistic prudence, i.e., to prudence unrestrained by altruism. It is just this prudential aspect of both utilitarianism and egoism that is rejected by the theory of living only for the moment and hence its name, "aprudentialism." Strictly, the term "aprudentialism" refers to any theory which prescribes the maximization of value (or minimization of disvalue) for a period of less than one's entire life—e.g., for one's old age only, or only for the next ten years. But here I use the term as an abbreviation for "aprudentialism of the present moment."

Utilitarianism and ethical egoism differ from each other on the question of whose good is to be taken into account. According to utilitarianism, of course, "everyone counts for one"; according to

ethical egoism, only one person counts, the agent. On this issue, aprudentialism can be construed along either utilitarian or egoistic lines. An aprudentialist can hold either that one ought to promote the present well-being of everyone, or that one ought to promote only one's own present well-being. To put the point another way, an aprudentialist preference ordering may include only egoistic values or both egoistic and altruistic values (or, a possible but rare case, only altruistic values). Thus, aprudentialism differs from both utilitarianism and egoism in rejecting the view that every time counts equally; but it is compatible with either utilitarianism or egoism on the counting of persons.

The principle of prudence—that one should seek to maximize one's total benefits without regard to when they occur—sometimes requires one to discount future benefits because they are less certain than present ones. If sacrificing now has only a small chance of paying off later, then, unless the sacrifice is very small and the payoff very large, prudence counsels against it. In some circumstances then, the act prescribed by prudence is the same as that prescribed by aprudentialism. But the reasons for the prescriptions are different. Prudence discounts future benefits in proportion to uncertainty about whether they will occur; aprudentialism holds that the future has no value not because it is uncertain, but simply because it is future.

The aprudentialist holds that one ought not to sacrifice present benefits for future ones, no matter how much greater those future benefits are. But if an aprudentialist does not value the future at all, how can we speak of "future benefits" in the context of the theory? We do so by assuming that all the events in question are to occur in the present. Thus, when we say that an aprudentialist chooses a present event having lesser value over a future event having greater value, we mean that if both events were alternatives for "now," the latter would be preferred; if there were no temporal difference between the events, the one which is in fact future would be preferred to the one which is in fact present. But since there is temporal difference, the future event, like all future events regardless of the values they would have if they were present, has no value at all for the aprudentialist.

## II

Let us turn now to a more detailed discussion of aprudentialism. The aprudential principle is, of course, a principle for the evaluation of acts, i.e., it provides a method for determining which of a set of alternative acts a subject ought to perform. The first step in applying the principle then is to identify the set of alternatives from which the act to be performed is to be selected. First, these alternatives must be acts which the subject in question has both the ability and the opportunity to perform—acts which the subject can perform. Further, these acts must be alternatives for a particular time. Aprudentialism does not deal directly with a dilemma like "Would it be better to visit grandmother today or to visit her tomorrow?" The alternatives considered in aprudentialism are not times a given act can be performed but rather acts which can be performed at a given time. So the question about grandmother must be put: "Should I visit grandmother today or (say) go swimming today?" But the aprudentialist is concerned not with days but with moments. Suppose I am driving a car at an intersection where left leads to grandmother's and right to the beach; the issue then is "Left or right?"—that is, what ought I to do at this moment? Thus, in determining what a subject ought to do, aprudentialism focuses on a set of acts each of which the subject can undertake at a given moment. To capture this notion I adopt, for the ideal version of the theory, the concept of a durationless point in time  $t$ . The set of alternative acts of concern to ideal aprudentialism then consists of just those acts which a subject  $S$  can perform at  $t$ .

The object of aprudentialism is to maximize present intrinsic value (or minimize present intrinsic disvalue). The bearers of intrinsic value and disvalue are events, including acts.<sup>3</sup> Thus, aprudentialism focuses on all those acts which  $S$  can perform at  $t$ , together with all events occurring at  $t$  which are brought about by such acts. An example of the latter is a light's going on when there is no delay in the mechanism;  $S$  pushes a button at  $t$ , causing the simultaneous event of the light's going on. I refer to these events—acts  $S$  can perform at  $t$  and simultaneous events brought about by such acts—as events  $S$  can bring about at  $t$ . In ideal aprudentialism, events, like points in time, are construed as having no durations.

<sup>3</sup> States of affairs are also bearers of intrinsic value and disvalue. One may value for its own sake the silence of the morning, or being healthy. For our purposes we may construe states existing at  $t$  as events occurring at  $t$ .

In making an aprudential decision we assume a preference ordering for *S* over events which *S* can bring about at *t*. This ordering is in terms of intrinsic values only. Extrinsic, or instrumental, values are not taken into account. Extrinsic value is of course value something has by virtue of its relation to something else. This relation may be empirical (e.g., watering the plants causes them to bloom) or logical (e.g., chewing is part of eating). Intrinsic value, in contrast, is value something has independently of its relation to anything else. An aprudential preference ordering then is in terms of intrinsic value only. This restriction accords with our intuitions about what it is to live only for the moment; an aprudentialist values events only for their own sakes, never because of their relations to other events.<sup>4</sup>

This relativization of intrinsic value allows us to eliminate from the theory the concept of intrinsic disvalue and also that of an event's having neither intrinsic value nor intrinsic disvalue. From an "absolute" point of view, *S* may judge that walking beside a swimming pool fully clothed has intrinsic disvalue. On our construal, *S* intrinsically prefers walking beside the pool to walking into it. Thus, all events which *S* can bring about at *t* can be included in a single preference ordering.

On the ideal construal of the theory, this preference ordering is conceived of as existing only at the durationless point in time *t*. The reason for this stipulation is that at each point in time the alternatives open to *S* are different. If I have just moved my arm, it is not now open to me to move the same arm in the same direction through the same part of space. Similarly, events occurring at *t* create new conditions in the moment immediately following *t*. Hence, we suppose that each preference ordering for *S* exists at a unique point in time and that for each successive point there is a new ordering.

But how are the events referred to in an aprudentialist preference ordering to be described? Suppose that one way of participating in a party is by dancing a jig, and that one way of dancing a jig is to go hop step, hop step, hop step, and so on. Does it matter whether we say that *S* prefers participating in the party or that *S* prefers dancing a jig or that *S* prefers going hop step, hop step, hop step? The answer is that it does matter, for we are concerned only with *S*'s intrinsic values. We must

determine then what it is that *S* prefers not because of its relation to anything else, but for its own sake. Is what has the most value for *S* participating in the party—never mind how? Or is it dancing, a jig—regardless of whether dancing counts as participating, and regardless of the way the dance is done? Or, is what *S* most prefers for its own sake going hop step, hop step, hop step—independently of whether these movements count as dancing a jig or as participating in a party? The answers to questions like these determine the way in which *S*'s intrinsic preferences are described. In stating *S*'s overriding intrinsic preference then, we use that description which mentions just those features by virtue of which *S* intrinsically prefers the event in question; we omit mention of that which *S* values only because of its relation to something else. Thus, if what *S* most prefers for its own sake is dancing a jig, without regard for the way the jig is done or for the jig as a way of participating in the party, we say simply that *S* most prefers dancing a jig. On the other hand, if *S* intrinsically prefers dancing, say, the traditional version (hop step, hop step) to the Balanchine variation (pirouette, pirouette, pirouette), we say that *S* most prefers dancing a traditional jig, or dancing a jig *by* going hop step, hop step, hop step.

Given *S*'s overriding intrinsic preference at *t* then, we formulate a prescription directing *S* to perform at *t* that act which is or brings about the preferred event. Such a prescription has the form "*S* aprudentially ought to do act *A* at *t*"; its justification is of the form "because *S* most prefers, for its own sake, event *A* to all alternatives which *S* can bring about at *t*." Ideally, such a prescription is formulated for each succeeding point in time.

What I have described so far is the ideal or limiting version of the theory of aprudentialism. Clearly, if the theory is to have practical application there must be a way of making prescriptions which mention not durationless points, but periods of time. It will not do, however, to select some duration—e.g., ten minutes, ten seconds—and label it "now," for there is no non-arbitrary way to decide how long "now" should be. Hence, the applied version of aprudentialism presupposes the ideal version and retains the identification of "now" with durationless points, but allows the formulation of preference statements and prescrip-

<sup>4</sup> Criteria of intrinsic value are not given by the theory. Aprudentialism resembles utilitarianism and egoism in requiring supplementation by a theory of value.



tions in terms of clock time. The relation of the ideal and the applied versions is analogous to the relation of ideal and applied geometry.<sup>5</sup>

In applying aprudentialism, we say, for example, that of all the events *S* can bring about for the hour between three and four o'clock today, *S* most prefers event *E* (where *E* is continuous at least throughout the duration mentioned). Such a preference statement is to be understood as implying that there is no time throughout the period in question when there is an event other than *E* which *S* can bring about and which *S* intrinsically prefers to *E*. That is, while in applying aprudentialism we speak of durations rather than of durationless points, the preference specified for a duration is a preference over all alternatives open to *S* throughout that duration, not merely a preference among the alternatives for the initial point of the duration.

But how do we determine the period of time to be considered? The initial point of the period is of course the time about which the aprudential question is asked. As we have seen, a question of the form "What, according to aprudentialism, ought *S* to do?" must always be qualified by mention of a particular time—e.g., "What ought *S* do right now?" or "What ought *S* do at 3:15 p.m. today?" The time specified, of course, is not given by the theory; it depends on the factors which create an interest in answering the question. The final point of the duration is not given by the theory either. It depends largely on our ability to predict. I may know enough about *S* and about the alternatives open to *S* for the next hour to say with confidence that *S* aprudentially ought, say, to dance a jig for 10 minutes and then to drink with friends for 50 minutes. But if I am in doubt about the alternatives at the end of the hour or about *S*'s preferences then, I have no firm basis for a longer-term prescription. Hence, the duration mentioned in a statement of preference and in a prescription based on it is, in a sense, arbitrary. Its beginning may be any point we choose; its termination, if our preference statement is to be accurate and so our prescription well-founded, should be no farther in the future than our foresight allows us confidently to predict.

In applying the theory of aprudentialism then,

we determine *S*'s overriding and intrinsic preferences throughout some duration about which we can make predictions with at least moderate confidence, and formulate a prescription directing *S* to bring about that event or series of events which *S* most prefers throughout the period in question.

One who prefers the present to the future and so chooses to act on such prescriptions is an aprudentialist, i.e., is one who chooses to live only for the moment.

### III

I turn now to arguments for the claim that aprudentialism is irrational. These arguments purport to show that preferring one time to another only because they are different times forces one to reject certain elements of our shared conceptual scheme—viz., the principle of universalizability, and the concept of a person.<sup>6</sup> The concept of rationality involved here is not that according to which to be rational is to choose the most efficient means to one's ends. Clearly, if one prefers to seek the greatest possible total of happiness throughout one's life it is, in the "efficient means" sense, irrational to adopt aprudentialism, just as, if one's ultimate end is the pleasure of the moment, it is in the same sense irrational to adopt prudence. But what is at issue here is not the relation of means and ends; it is, rather, whether aprudentialism requires the rejection of certain well-entrenched concepts.

The first step of the attack on the rationality of aprudentialism is that to prefer one event to another merely on the basis of time is irrational because it violates the principle of universalizability. If the only difference between two events is that one is present and the other future, there is, according to this argument, nothing to choose between them—the difference is only apparent. Every present event is at some time future and every future event is at some time present. Of two events identical except for time then, both are—tenselessly—both present and future. They are wholly identical. Thus, one can have no reason for preferring one to the other; to do so is irrational.

The aprudentialist reply to this argument points out that there are two ways of construing the

<sup>5</sup> It may be suggested that aprudentialism should be construed in terms of the principle that there is to be an increasing discount on the value of events the farther they are in the future. On this view, when a future event is of much greater intrinsic value than a present one, one may be required to sacrifice the present for the sake of the future. I do not adopt this suggestion because I am concerned to explicate a position according to which one should never sacrifice present for future.

<sup>6</sup> The arguments developed in this section are based on the work of Nagel and Rawls cited in n. 2.

choice between present and future. On one hand, the alternatives may be described tenselessly, as they are in the argument against aprudentialism. On this kind of description, time preference generates contradiction because a characteristic which is a reason I have for preferring a certain event is a characteristic of every event. (I prefer *X* to *Y* only because *X* is present; but there is also a time when *Y* is present; hence I must also prefer *Y*, which is contradictory.) On the other hand, the choice between present and future may be described in relation to particular times. On this kind of description there is no contradiction because all events do not have the same temporal characteristics. (I prefer *X* to *Y* only because *X* is present; there will be a time when *Y* is present, but that time is not now; so there is no contradiction.) The objector, struck by the fact that every event exists at some time, uses verbs tenselessly. The aprudentialist, struck by the difference between what exists now and what will exist, formulates the issue in tensed statements. What is needed is argument to show that the tenseless construal is required.

The argument addressed to this point holds that to reject the tenseless formulation is to reject the concept of oneself as an individual persisting through time. To be a person is to be temporally extended, it is to have a past and future as well as a present. But to be an aprudentialist, to care about only the present, is to be "dissociated" from one's future self—it is to act as though one had no future. But more than that, to refuse to take prudential considerations into account is also to be alienated from one's present self. For in rejecting prudence one does not treat one's present self as what it is, viz., a stage in an ongoing being, but rather, one makes decisions as though one were a temporary or even momentary consciousness. Hence, to reject prudence is both to fail to identify with oneself as a continuing individual and also to fail to identify with one's true present self.

"Identification" and "dissociation" may suggest that this is a psychological argument claiming that for the sake of one's long-term mental health one ought to be prudent. But there is no mere question-begging here. The claim is, rather, that aprudentialism precludes a view we all have of ourselves: that we are temporally extended beings, that the present moment is just one of a series of moments which are all equally real parts of our lives.

In reply, it must be shown that one can believe that one has a future (i.e., that one can have a concept of a self as a persisting being and the belief that one is such a self) even though one does not value that future. Ordinarily, of course, it is easy enough to give an account of what it is to believe in the existence of something one does not value. For example, I may be wholly indifferent to a certain coffee cup, but if it comes flying at my head I duck, thereby demonstrating my belief in its existence. In the case of the aprudentialist and the future however, no such demonstration is possible, since as a matter of logic the future cannot influence anything the aprudentialist cares about. It is logically impossible for a future event to have causal efficacy in the present. Hence, it will not do to argue that just as an ethical egoist may consistently value the well-being of another person not as an end in itself but as a means to the egoist's own well-being, so an aprudentialist may value the future as a means only. Other people can affect the egoist's well-being, but future events cannot affect the present. We cannot show that the aprudentialist believes in a continuing self by arguing that future states of the self may be valued as means.

Another attempt to demonstrate the aprudentialist's belief in an enduring self relies on the fact that one's present attitude toward the future—e.g., worrying about it, or happily looking forward to it—affects one's present well-being. It would seem that aprudentialism directs one to eliminate worry and to promote pleasant anticipation in the present and that therefore aprudentialism is consistent with having a concept of one's future. But to have a future-directed attitude of worry or anticipation is to care about the future. We have already seen that it is a mistake to care about the future as a means of improving the present. And the aprudentialist does not care about future events as ends in themselves since to do so is inconsistent with aprudentialism. Hence, it seems that aprudentialism precludes future-directed attitudes and activities. But if interest in the future has no place in the aprudentialist's life, the aprudentialist has no use for a concept of a future self.

The above argument assumes that the aprudentialist is "pure" in the sense of actually valuing only what is valuable according to the theory. If this assumption is false it is easy to show that aprudentialism and the concept of a temporally extended self are compatible. Consider then that one's values may include not only what one chooses

to value, but also what one values without choosing. If we ourselves were to adopt aprudentialism for example, we would probably find that we still cared about our futures, whether we wanted to or not. If such caring is beyond one's immediate control, it is a "given" in relation to aprudentialism—something to be accepted and dealt with in accordance with the theory. Suppose I find that I feel very anxious about whether I shall have any money next month, and that I am not able to eliminate this anxiety by willing it gone. In this case it may be that the most satisfying thing I can do right now is to see about getting some money for next month. If so, then this is what I, as a good aprudentialist, should do. Aprudentialism prescribes future-directed acts where they are superior to any available alternatives in increasing the value of the present. (It may be suggested that one who is committed to aprudentialism but still values the future should undertake to overcome this attachment, to free the self from this unchosen concern. But this is so only if the events constituting such a project of personality change are intrinsically preferred, at the times they occur, to all alternatives for those times.) It is clear then that if people who are deeply habituated to caring about the future adopt aprudentialism, they will frequently be required, on aprudentialist grounds, to perform acts whose intention is to alter the future. Such an act requires the agent to have a concept of a future self; aprudentialism, then, does not require one to reject such a concept.

But to show that an impure aprudentialist has a concept of self as continuous is not of course to give an adequate reply to the argument that aprudentialism requires one to conceive of selves as momentary. What is needed is argument to show that a *pure* aprudentialist views selves as continuous. Notice then that aprudentialism is not something that happens to one (infants are not aprudentialists), but something one chooses. The aprudentialist is by definition one who regularly chooses to sacrifice the future for the sake of the present; it is this pattern of choice that makes one an aprudentialist. But if one chooses among a set of alternatives, one has concepts of those alternatives. The aprudentialist chooses the well-being of a person now over the well-being of that same person later. Thus, the aprudentialist employs a concept of a person as continuous in making the commitment to aprudentialism. Aprudentialism does not preclude conceiving of selves as persisting through time; on the contrary, in

order to choose aprudentialism one must have and use such a concept.

It may be objected that a pure aprudentialist, after making an original commitment to the position, would no longer have use for a concept of a continuing self since every moment of the aprudentialist's time should be taken up with something else. But to be a subscriber to a normative theory it is not enough simply to act as prescribed by the theory. One who is a utilitarian, for example, is one who on some occasions makes judgments by appeal to utilitarianism. Similarly, one who is an aprudentialist sometimes judges in accordance with that principle. Thus, an aprudentialist who is "pure" always prefers the present and sometimes prefers, for the present, making judgments about what ought to be done. The pure aprudentialist periodically renews the commitment to aprudentialism and in doing so employs a concept of a temporally extended self.

The argument to which we are responding holds that to reject prudence is to fail to view oneself as a continuing being and to fail to see one's present self as a stage in that continuity. The reply is that in opting for an aprudentialism which is not wholly altruistic, i.e., which includes a preference for one's own present satisfaction over one's own future satisfaction, one necessarily has and uses both a concept of one's self as persisting through time and a concept of one's present self as a part of that continuing self. It is not the case then that to reject, for the purpose of choosing among events, the tenseless construal of such choices is to adopt some anomalous concept of one's self as a momentary existence. An aprudentialist who deliberately chooses the present well-being of the self over the future well-being of that same self is not adopting a non-standard view of the self, but only a non-prudential way of evaluating events.

I conclude, then, that aprudentialism has a place among those theories of conduct which account for normative arguments about what ought to be done. That is, aprudentialism has a role as a descriptive theory. I have not argued that it ought to be adopted as a guide to action and the evaluation of acts. In this regard I offer only a brief suggestion for the interpretation of aprudentialism which may help to show why it sometimes is adopted as a normative principle. The aprudentialist's attitude may be in part like that of the moralist who holds that since persons are intrinsically valuable, deliberately to harm one for the sake of another is intolerable, even if the benefits

to be produced are greater than the harm. The extreme pacifist will not kill a person even if this is the only way to save the lives of many others. Similarly, the aprudentialist may feel it an unacceptable violation of the value and integrity of the present to spoil it deliberately, regardless of whether doing so will produce greater benefits later on. Or to put the point another way: For the aprudentialist, each moment is to be treated as an end, never as a means only.

*Washington University*

*Received April 6, 1973*

## VI. RITUALS AND "OUGHT"

STEPHEN SCOTT

### I

SOME people, like the commandant of Auschwitz, ought to shirk their duties, and anyone should default an obligation to behead John the Baptist. It is nevertheless consistent with these cases that promises, contracts, and stations, by creating duties and obligations, sometimes determine what an agent ought to do; a familiar account holds that they do so unless overridden, that it is valid to argue:

- (1) *A* has a duty (an obligation) to do *x*.
- (2) *A*'s duty (obligation) to do *x* is not overridden.
- (3) So, *A* ought to do *x*.<sup>1</sup>

Two duties conflict if not both can be performed, and one is non-overridden merely if its owner has no conflicting duty or obligation that he ought to do instead; it is not part of the meaning of "non-overridden" that he ought to do the duty or anything else.

The belief that "(1) (2) so (3)" is valid can survive surface change without metamorphosis of its core motivation—that the only evaluative claims needed to entail (3) are about duties and obligations. A claim about some complex relation between the action *x* and the property *being a duty*—say, that *x* is more of a duty than any conflicting one—can take the place of (1), or of both (1) and (2). Also, all that is essential about (2) is that it is non-evaluative or, failing that, its support

lies in some relation between *x* and *being a duty*. The soggy concept of an "evaluative" sentence will not hold up long; let us speak instead of an "ought-generating" sentence, one that entails any instantiation or closure of the form:

- (4) — ought to do —.

The core of the thought that "(1) (2) so (3)" is valid—what I shall call the "prima facie view"—is then:

- (3) follows from (a) some sentence(s) asserting a relation between *x* and *being a duty* (*being an obligation*) and, if needed, (b) some analytic sentence(s) and (c) some contingent sentence(s) that either is not ought-generating or else is entailed by sentences of type (a).

The first purpose of the paper is to show that this core requirement is unsatisfiable, the second to describe the modest support that (1) does give to (3).

### II

There are, as always, preliminaries. The initial one is to note that "duty" and "obligation" have two senses, general and restricted, the latter being the normal (unsignalled) one in this paper. In the former, they are loose synonyms for each other and for "things one ought to do," but in the tight sense non-synonymous, duties being actions required to accomplish the tasks of stations, obligations to keep one's word.<sup>2</sup> The proper senses are then related to

<sup>1</sup> "(1) (2) so (3)" is frequent as the last move of longer arguments. J. L. Austin sets out quickly an argument that moves from "performative" to "contract" to "obligation" to "ought." (*How to Do Things With Words* [Oxford, 1962], p. 53), John Searle (in "How to Derive 'Ought' from 'Is'," *The Philosophical Review*, vol. 73 [1964], pp. 43–58, and later in *Speech Acts: An Essay in the Philosophy of Language* [Cambridge, 1969], pp. 175–198) fills in the line Austin sketched, using "promise" rather than "contract." The Austin-Searle line, if valid at all, holds equally for the inference from "station" to "duty" to "ought."

The original of the line is from Hobbes, *Leviathan*, Pt. I, ch. 14: "Right is layd aside, either by simply Renouncing it; or by Transferring it to another. . . . And when a man hath in either manner abandoned, or granted away his Right; then is he said to be OBLIGED, or BOUND . . . : and that he Ought, and it is his duty, not to make voyd that voluntary act of his own. . . . The way by which a man either simply Renounceth, or Transferreth his Right, is a Declaration, or Signification, by some voluntary and sufficient signe, or signes. . . . And these Signes are either Words onely, or Actions onely . . . or both. . . ."

"The mutuall transferring of Right, is that which men call CONTRACT."

Hobbes is correct in noting that actions can effect a contract, although I later speak of contracting as "saying something"; this is an indifferent carelessness, it being ritualistic that such and such actions effect a contract.

<sup>2</sup> The distinction between "duty" and "obligation" is discussed by R. B. Brandt in "The Concepts of Obligation and Duty" (*Mind*, vol. 78 [1964], pp. 374–393) and perhaps more nicely by James K. Mish'alani in "'Duty', 'Obligation' and 'Ought'" (*Analysis*, vol. 30 [1969], pp. 33–40). Mish'alani also sets out the two senses of each.

Since the repetition of "duties and obligations" is wearisome, I often speak only of duties, the extension being easy.

the general as to "ought." Duties of beneficence and of gratitude, not being created by stations, are general ones if they exist at all, and "a duty to keep promises" either confuses the tight senses or says loosely that promises ought to be kept.<sup>3</sup> "(1) (2) so (3)" is trivially valid for general duties—(2) is in fact otiose—and, since the concept of "general duty" contains not the least hint of what counts as one, proves only that a person ought to do what he ought to do. The argument is significant about special duties, being the hinge of any derivation of an "ought" claim from the description of a station.

### III

The second preliminary is to determine the relations between claims of the forms, "*A* ought to do *x*" and "*A* ought not to do *x*." It is in the nature of the humdrum that they can be true simultaneously *if* moral rules are formulated without "*ceteris paribus*" clauses or *if* detachment is applicable to hypothetical imperatives. For instance, these true hypotheticals have simultaneously satisfiable antecedents; given the applicability of detachment, their consequents can be true together: "If an officer wants honor, he ought to be first out of the trench; but he ought to dally, if he wants to survive the war." It is possible then that "ought" has a sense defined by the joint satis-

fiability of "*A* ought to do *x*" and "*A* ought not to do *x*."<sup>4</sup>

In a second sense, "*A* ought to do *x*" is not consistent with its apparent contrary. For it is often a judgment stating the *outcome* of deliberation. An officer, having detached the relevant hypotheticals and noted the dismaying result, still deliberates about what he ought to do—a deliberation possible only if he treats "ought" as having two senses; otherwise, he already knows what he ought to do, and too much about it. So far as he can act on it, the judgment that he ought to do something closes the possibility that he ought not to do it.

I shall call the first (perhaps non-existent) sense "derivative" and signal it by writing "ought<sub>d</sub>"; the second, signalled by "ought<sub>c</sub>" is the categorical sense. The features of "ought" that, under certain assumptions, lead to the conclusion that it has two senses—one satisfiable together with its contrary and one not—are familiar in philosophical literature.<sup>5</sup>

Duties are straightforwardly, insignificantly, connected to "ought<sub>d</sub>." It seems analytic that the duty of a station is what one ought to do as occupying the station; that, therefore, having a duty to do *x* entails that one ought to do *x*, even without the "non-overridden" rider. "Ought" here is "ought<sub>d</sub>." For one can have a duty to do *x* and another that precludes doing it. If having the first entails that one ought (in some sense) to do *x*, then, equally, having the second entails that one

<sup>3</sup> See Sir David Ross, *The Right and The Good* (Oxford, 1930), pp. 19–21. Ross holds what in this paper is called the *prima facie* position—not by his claim that *prima facie* duties are duties unless overridden, this being an obvious analyticity on his definitions, but by his further claim that an agent has a *prima facie* duty to keep his promises.

<sup>4</sup> "Possible"—that is, "in the absence of conclusive considerations to show otherwise." The formulation of moral rules is clouded beyond discussion. There are at most three ways to deny the applicability of detachment to a hypothetical imperative, that is, to the form:

(4.a) If *A* wants to get *y*, *A* ought to do *x*.

The first is to hold that, appearances to the contrary, (4.a) is not a conditional; the most plausible alternative here is that "ought" is the principal sign, the conditional falling within its scope. The second is to hold that "*A* wants to get *y*" does not instantiate the antecedent of (4.a). It appears that both Wilfrid Sellars (*Science and Metaphysics* [New York, 1968], pp. 191–192) and R. M. Hare ("Wanting: Some Pitfalls," in *Agent, Action and Reason*, ed. by Robert Binkley, et al. [Toronto, 1971], pp. 85–86) take this line, holding that the freestanding antecedent of (4.a) means exactly what it says but, as nested in (4.a), means something different. The third is to hold that "*A* ought to do *x*" does not instantiate the consequent of (4.a) because "ought" is equivocal. Kant makes this move, labeling "ought" in "*A* ought to do *x*" categorical (a different sense of "categorical," by virtue of being stronger, from the sense used in this paper) and in (4.a) hypothetical. So far as I know, there is no proof that any of the above analyses is correct.

<sup>5</sup> Brandt (*op. cit.*, p. 378) notes the use of "ought" sentences to close deliberation, the most important feature for establishing that "ought" is sometimes categorical. Brandt believes also that there can be conflicting "ought"s, as does W. D. Hudson, "The 'Is-Ought' Controversy," *The Is/Ought Question* (London, 1969), pp. 172–173. Roger Trigg ("Moral Conflict," *Mind*, vol. 80 [1971], pp. 41–55) notes that "ought" is sometimes used to describe the clash of moral demands, sometimes to express a final decision about a course of action (p. 52), and draws the conclusion that "ought" has two senses (p. 54). See also Hector-Neri Castaneda (pp. 266–271) on the qualified and unqualified uses of normative words ("Imperatives, Decisions, and 'Oughts'," *Morality and the Language of Conduct* [Detroit, 1965], pp. 219–299).

The distinction between "categorical" and "non-categorical" does not necessarily coincide with that between "moral" and "non-moral."

ought not (in the same sense) to do  $x$ ; all that follows simply from "duty" is "ought<sub>d</sub>." Yet only "ought<sub>c</sub>" holds any interest, since the live problem concerns the extent to which having a duty justifies doing it. That duties are *somehow* reasons isolates nothing special about them; that the moon is empty of cheese is a knockdown reason for the Borden Dairy not to build spaceships. It is quite unobvious—the officer's muddle is one of a plethora of confirming instances—in what way and to what extent "ought<sub>d</sub>" claims are reasons for acting. The *prima facie* view would not explain how duties justified actions if it related them merely to "ought<sub>d</sub>," just as it would have no consequences about *what* actions could be justified if it related "ought<sub>c</sub>" only to general duties.

## IV

If the argument "(1) (2) so (3)" is construed, however, as about "duty" and "ought<sub>c</sub>," it is invalid, since an agent's non-overridden duties, like his duties simply, can conflict. Any "ought" consequent upon one of them is then satisfiable with "ought not," and, therefore, is not "ought<sub>c</sub>."

Stations are consequences of the social division of labor; their duties are what are required to accomplish the tasks that division creates. Stations and duties are artifacts; like games and rules, they can be made as human beings choose, even if familiarity seems to give some—of a parent, of a spouse, of a citizen—the fixity of natural objects. (It is not the biological relation of *being a parent* that creates duties but the social [and perhaps legal] relation.)

Shortsightedly or malevolently, human beings can make conflicting, non-overridden duties. Inconsistent courses of action are non-overridden if both are equally permissible or forbidden; so also, nothing in the concept of "duty" entails that only one from a conflicting set is non-over-

ridden.<sup>6</sup> An evil bureaucrat, in need of a sure device for punishing his enemies and rewarding his friends, can create such duties; his friends will then be found always to have done theirs (whatever they do), his enemies always to have failed theirs (again, whatever they do).<sup>7</sup> Someone might object that the evil bureaucrat, in so acting, does not impose real duties, that he does not meet the conditions for successful imposition, that he has reneged. The objection lacks cynicism, as it would to suppose that a sergeant, demanding a day's work in five minutes, only flubs his attempt to give a command, when he actually creates a dire impasse. Most stations do contain regulations to resolve the unhappy results of conflicting duties, but these regulations are no more than conventions; they are not conceptual limits. Something can be a station and lack them, as a game might lack provisions to resolve inconsistent rules. The duties a person has, whether of the same or different stations, are duties, even if conflicting ones cannot be assigned priorities.

The same point applies to obligations. Anyone can make contradictory contracts, and it is merely contingent that there are devices to remove contractual blocks. Accordingly, that an agent has a non-overridden obligation to do an action does not entail that he ought<sub>c</sub> to do it, for the same reason that his having an obligation (without a qualifier) does not.

## V

Given that "(1) (2) so (3)" fails, it may seem that the *prima facie* view can be patched by replacing premise (2) with:

- (5)  $A$ 's duty to do  $x$  overrides  $A$ 's duties and obligations that conflict with  $A$ 's doing  $x$ .

Since two duties cannot each override the other—from the meaning of "override"—the argument

<sup>6</sup> There is perhaps an inclination to think that conflicting, non-overridden duties are not really duties, since, if such duties yield any "ought<sub>c</sub>" claims, they yield that one ought<sub>c</sub> to do both or neither of contradictory actions. To accept the inclination is to assume part of what is disputed, that something is a duty only if it entails an "ought<sub>c</sub>" claim; it is to assume that the *prima facie* view states a rock bottom feature of the concept of "duty." If that is to be given, then what is assumed in the paper, that stations create duties, is no longer obvious; the problems that arise now in linking duties to "ought<sub>c</sub>" arise then in deriving duties from the regulations governing stations; it is the same difficulty, only in a different place. The inclination is really no more than a slip from "duty" to "general duty."

<sup>7</sup> The Inquisition held that a person confessing under torture to witchcraft was guilty, by virtue of confessing; that a person refusing to confess, even under torture, was also guilty, since the strength to withstand such pain could come only from the devil. If tortured, guilty no matter what. Henry Charles Lea, *The Inquisition in the Middle Ages*, abridged by Margaret Nicholson (New York, 1961), pp. 816–817.

The Inquisition did not submit its friends to torture but, by a similar delicate line of reasoning, would have always found them innocent.

form "(1) (5) so (3)" escapes the difficulty fatal to the original. Even so, a matching difficulty follows upon the substitution of (5) for (2).

(5) is equivalent to:

(6) *A* ought<sub>c</sub> to do his duty to do *x* rather than any conflicting duty or obligation.

Since (6) either asserts that one ought<sub>c</sub> to do *x* to the exclusion of conflicting duties and obligations or ranks doing *x* above opposing alternatives without asserting that one ought<sub>c</sub> to do any of them, (6) is at most ambiguous between:

(7) *A* ought<sub>c</sub> to do his duty to do *x* and not any conflicting duty or obligation.

and:

(8) If *A* ought<sub>c</sub> to do any of his duties and obligations, then, assuming that he has a duty to do *x*, he ought<sub>c</sub> to do *x* and not any conflicting duty or obligation.

Now, a sentence of form (7) is an ought-generating claim and, there being false instances of (7), is not a conceptual truth about "duty"; so, if (5) is equivalent to (7), the revised argument "(1) (5) so (3)" fits the core of the *prima facie* only if the grounds for a sentence of form (7) lie in the relation of *x* to *being a duty*. In turn, if there is any relation between *x* and *being a duty* sufficient to generate (7), it would seem to be either:

(9) *A* has a duty to do his duty to do *x* and not any conflicting duty or obligation.

or

(10) *A*'s duty to do *x* is more of a duty than is any conflicting duty (and more of a duty than any conflicting obligation is of an obligation).

One can have, as (9) alleges, a duty to do a certain duty, since a station can include an overriding device. A ship's captain has both a duty to safeguard the cargo and a duty to preserve the passengers; he might have a further duty to do the latter rather than the former when pressed. This much conceded, (9) is insufficient to entail (7), (9) being satisfiable simultaneously with its contrary:

(11) *A* has a duty not to do his duty to do *x* and to do instead some conflicting duty or obligation.

(The evil bureaucrat who provides in one regulation that a ship's captain should first preserve the passengers can stipulate in another that cargo

comes first—even second-order regulations can be inconsistent.) Since (7), because of "ought<sub>c</sub>," is not satisfiable with its contrary, (9) does not entail (7).

Neither does (7) follow from (10), although there are some true sentences of form (10). For instance, suppose that a general has duties to prepare defenses against invasions and to minimize military transgressions of civilian rights; then for him the former is more of a duty than the latter, since doing it is, and doing the latter is not, necessary to perform the function of a general. It does not follow, however, that the former overrides the latter in cases of conflict; considerations irrelevant to being a soldier—perhaps a citizen's duty to uphold the constitution—can reverse the ordering. That one action is more of a duty than another provides an internal ranking of the duties of a single station; it would determine that one ought<sub>c</sub> to do one duty rather than another only if it also ranked the duties of different stations. That it does not is clear in examples. The duties of a parent to care for his children and of a night-watchman to guard the plant are equally duties, since guarding the plant is as essential to a night-watchman's function as is caring for his children to a parent's. The duty of a parent to educate his children is less of one than the guard duty of a nightwatchman, it being possible to care for children without educating them. Yet it generally happens, if conflict arises, that both parental duties override the nightwatchman's. This would not be possible if (10) entailed (7).

To summarize the argument so far: If one construes (5) as equivalent to (7), then the choice of (5) to bridge the gap between (1) and (3) violates the core of the *prima facie* view. For a sentence of form (7) is ought-generating, is not a consequence of the meaning of "duty" and, (7) being entailed by neither (9) nor (10), is not grounded in any relation between an action and *being a duty*.

Suppose, then, that one takes the other alternative and construes (5) as equivalent to (8). (8) and (1) entail (3) only in conjunction with the further assumption:

(12) *A* ought<sub>c</sub> to do at least one of his (conflicting) duties and obligations rather than neglect them all.

(12) is ought-generating, and, not mentioning any, its truth does not hinge on the particular duties *A* has; accordingly, "(1) (8) so (3)" is consistent with the *prima facie* core only if (12)



states a conceptual feature of duties and obligations.

It does not do so. It entails:

- (13) *A* ought<sub>c</sub> to do only one of his conflicting duties and obligations.<sup>8</sup>

For the argument form, "*A* ought<sub>c</sub> to do *x*; *A*'s doing *y* prevents him from doing *x*; so, *A* ought<sub>c</sub> not to do *y*," is valid. So, if an agent's duties conflict and he ought<sub>c</sub> to do more than one of them, then he also both ought<sub>c</sub> and ought<sub>c</sub> not to do one of them, which is impossible. Yet (13) does not follow from the meaning of "duty," since, as noted earlier, there can be impasses among duties; if the captain has inconsistent second-order duties, an examination of his duties does not rule out that he ought<sub>c</sub> and ought<sub>c</sub> not to do one of them. (13) being a consequence of (12) but not of the meaning of "duty" or "obligation," (12) itself is not an analytic truth about duties and obligations.

The construal of (5) as (8) requires the introduction of (12) into the premisses and, therefore, lapses from the prima facie view. Since (7) and (8) exhaust the possible interpretations of (5), since the construal of (5) as (7) also fails, "(1) (5) so (3)" itself violates the prima facie core.

## VI

The arguments in Sects. IV and V show cumulatively if drily that premisses whose sole ought-generating claims concern duties and obligations do not determine what someone—say, *A*<sub>1</sub>—with conflicting ones ought<sub>c</sub> to do. For *A*<sub>1</sub> ought<sub>c</sub> to do one only if he ought<sub>c</sub> to do it rather than the alternatives. The "only if" clause here either is or requires an ought-generating claim among the premisses of (3). Yet nothing in the properties *being a duty* and *being an obligation*, and nothing in any relation of these properties to particular actions, entail that *A*<sub>1</sub> ought<sub>c</sub> to do a given duty over its alternatives.

There perhaps remains a mild and last inclination to derive (3) from (1) plus:

- (14) *A*'s duty (obligation) to do *x* does not conflict with any of his other duties or obligations.

(14) also limps. A person can manufacture at will obligations inconsistent with a given obligation or duty. If he has a duty to do *x*, he can (seriously) promise not to do it; promising can be done when and as he wants. (He can make a vow if there is no one to receive a promise; a vow of poverty creates an obligation but not an obligation to anyone.) With the same ease one can conjure up conflicting duties. Appointing someone to a station consists in saying the right words in the right circumstances; what counts as "right" is stipulative initially and conventional later. For any task, then, one can devise a self-appointive station having that task as its duty.<sup>9</sup>

An agent's duty is non-conflicting only if he wants it to be so. Yet statements of duty, if they are reasons for "ought<sub>c</sub>" statements at all, are so independently of wanting them to be; otherwise, distaste would be a quick relief from the encumbrance of an unpleasant duty. Possession of a non-conflicting duty to do an action does not then entail that one ought<sub>c</sub> to do it; here also (3) is not derivable along the lines of the prima facie position.

## VII

An extension of the word "categorical," so far applied only to "ought," will bring to daylight the worm in the core position. Let *W* be a variable for any phrase that makes a sentence form of "*W* that *p*" or of "*W* to do *x*"; then *W* is non-categorical if the form "*W*(that *p* and not *p*)" or "*W*(to do *x* and not to do *x*)" is consistent, otherwise categorical. "*A* says that *p*" does not generally entail that *p* simply because "*A* says that..." is not, and "It is true that..." is,

<sup>8</sup> Let *x* and *y* be *A*'s conflicting duties and obligations. (13) does not have the force of the disjunction: *A* ought<sub>c</sub> to do *x* and not *y*, or *A* ought<sub>c</sub> to do *y* and not *x*. The disjunction does not follow from (12). Rather (13) has the force of: *A* ought<sub>c</sub> to do (*x* and not *y*) or (*y* and not *x*). To show that the former does not, but that the latter does, follow: Suppose that *A* ought<sub>c</sub> to go to the store and that the only two means are by car or on foot, neither preferable to the other. It does follow that *A* ought<sub>c</sub> to go either by car or on foot, but not that either he ought<sub>c</sub> to go by car or he ought<sub>c</sub> to go on foot—the disjuncts being exclusive and yet either way of going being acceptable.

<sup>9</sup> Not, of course, that one can enter any station by self-appointment but that, for any task, one can devise some station so entered. There are familiar stations of this sort. One can become a child's guardian, at least in some circumstances, without appointment by a third party and without the ward's consent. Keeping the nightwatch is a duty open to almost anyone interested, as are a host of others, all generally unpleasant, their possession accompanied by few (if any) rights.

categorical. The special cases in which the entailment (or, to hedge, the "guarantee") holds fit the same rule. Sometimes a person promises or contracts by saying he does; yet promising to do  $x$  and not to do  $x$  is as easy as saying that  $p$  and not  $p$ . In the special cases, if  $p$  is completed by an infinitive or by a "that" clause, the phrase governing the infinitive or the clause is non-categorical.<sup>10</sup>

All efforts to make the *prima facie* view work try either to dispense with categoricity in the conclusion (3) or to import it into the premisses. Dispensing is here decapitating, because only "ought<sub>c</sub>" conclusions justify actions; and importation either falls short—the qualifiers "non-overridden" and "non-conflicting" do not bring categoricity with them—or cheats—"overriding" slips in an ought-generating claim. Original sin does not vanish. If some ritual commands doing  $x$ , a second can forbid it, and the devices creating duties and obligations—swearing under oath, contracting, marrying, being naturalized, being ordained, and so on—are all rituals—to clothe them philosophically, things accomplished by performatives. It is conceptually possible for an agent to perform any two rituals, even if it should be posited in one that he is not to perform the other.<sup>11</sup> One can say anything, can stare in the face the rules about what one should say, and ignore them. The thin air that creates duties and obligations does not yield a categorical "ought." The principles determining what an agent ought<sub>c</sub> to do are perhaps self-legislated, as Kant thought, but this legislating is not merely uttering a performative.

### VIII

Yet it is clear that duties and obligations *support* "ought<sub>c</sub>" claims, even if only in conjunction with independent ought-generating sentences. Given this much, some further features about the relations among "duty," "obligation," and "ought<sub>c</sub>" are obvious.

First: Often one can determine that a certain action ought<sub>c</sub> to be done without determining by whom it ought<sub>c</sub> to be done. One knows that children ought<sub>c</sub> to be educated, but nothing in this knowledge determines that a given individual or a given group of individuals ought<sub>c</sub> to educate them. (It is false, of course, that everyone ought<sub>c</sub> to do it.) If there is to be a cocktail party, one knows that someone needs to provide drinks, someone clean house, someone invite guests, and so on, without knowing of any designated individual upon whom it falls to do one or more of these things. Now both contracts and stations are devices for assigning responsibility for actions to particular persons, and an agent, by giving his word or by assuming a position, takes something upon himself. That it is his (overriding) duty to do  $x$ , then, answers the question, "Who ought<sub>c</sub> to do  $x$ ?" given that someone ought<sub>c</sub>. Someone ought<sub>c</sub> to care for  $B$ 's cat in  $B$ 's absence; if  $A$  promises, then it is  $A$  who ought<sub>c</sub> to do so.

Second: One cannot always determine that someone ought<sub>c</sub> to do  $x$  independently of a particular person having a duty or an obligation to do  $x$ . In this circumstance, the role of duties differs from that of obligations. I handle duties first.

From the fact that someone ought<sub>c</sub> to educate children it does not follow that someone ought<sub>c</sub> to teach the eighth grade or beginning arithmetic or anything similar. Neither are the grounds for the latter of the same sort as for the former, it being reasonable that, in any society, someone ought<sub>c</sub> to do the former, it being at least unprovable that an elementary school system is the necessary condition. Usually an agent's duty is not simply to teach but to do some specific task. For stations create a division of labor; they subdivide a task into component tasks. One then establishes that  $A$  ought<sub>c</sub> to do his (overriding) duty to do  $x$  by establishing that his doing  $x$  is one component of task  $y$  that someone ought<sub>c</sub> to do.  $A$  ought<sub>c</sub> to do his duty to teach the eighth grade because doing

<sup>10</sup> Sometimes saying that  $p$  guarantees that  $p$ , and yet  $p$  contains no infinitive or "that" clause. To use Austin's examples, if  $A$  says that he marries  $B$ , or that he places a bet with  $B$ , or that he congratulates  $B$ , or that he christens the ship "Titanic," then sometimes he does the thing. (*How to Do Things*, *op. cit.*, pp. 5 and 40.) The remarks about non-categoricity can be extended to these cases. In some,  $p$  is synonymous with a sentence containing an infinitive or "that" clause; there the relevant phrase is non-categorical. For instance, to marry or to place a bet is simply to contract to do something, and " $A$  contracts to . . ." is non-categorical. In the others, an agent can do at once both the thing and its opposite; he can congratulate and express envy; he can christen the ship "Titanic" and "Queen Mary."

<sup>11</sup> As noted in Sect. IV, there are procedures to prevent the performance of conflicting rituals, but these procedures make nothing conceptually impossible. Perhaps, in the absence of divorce, only the first time a person marries counts—but it is only a convention that this is so, and human beings can choose to supplement it with a conflicting one. (Perhaps the supplemented ritual would not be called "marrying"—this also is a matter of choice.) "It is a convention that  $p$ " can be true together with "It is a convention that not  $p$ ."

so is one part (within the present division of labor) of educating children.

Contracts function not to create a division of labor but to secure the exchange of private property. One can often determine that a person has a right to make a certain kind of exchange, without determining that he has a right to make *this* exchange with *this* person. (In this context, one party has a right to make an exchange if the other party ought<sub>c</sub> to consummate the exchange.) For instance, only by exchanging money for commodities can one live and work within a system of private property that includes a highly articulated division of labor. Assuming that *A* lives in such a system and that he has a right to life, he has a right to exchange money for commodities; it does not follow that he has a right to buy commodity *C* from person *B*. Yet his right can be cashed only by making particular exchanges, it being by a contract then that he moves from a right to a type of exchange to a right to a particular exchange of that type. If *B* contracts to sell *C* to *A*, *A* has a right (if *B*'s obligation is overriding) to buy *C* from *B*—*B* ought<sub>c</sub> to sell, and, by a mirror argument, *A* ought<sub>c</sub> to buy. Where there is no right to the type of exchange, a contract counts for nothing.<sup>12</sup> For instance, if two people contract to turn somersaults on a street corner—nothing at all hinging on it, neither money nor expectations nor amusement—the contract is idle and does not establish that either ought<sub>c</sub> to turn somersaults. An agent can seriously take upon himself inane things; no overriding factors are then needed to allow default, a modicum of enlightenment sufficing. It seems silly that the part of him that acts should remain categorically bound to consistency with the part that talks. In a more serious case, a contract to buy another's life is void, no one having a right to buy that kind of thing.

Third: An agent ought<sub>c</sub> to do his duty only if it overrides conflicting duties and obligations. Otherwise, it can be true that someone ought<sub>c</sub>

to do *x*, that *A* has a duty to do it, and false that he ought<sub>c</sub> to do it. For that *A*'s duty is overridden entails only that he—not that everyone—ought<sub>c</sub> to do something different. (What is wrong with the commandant of Auschwitz doing his duty is not that it is overridden but simply that neither he nor anyone else ought<sub>c</sub> to do it, whether it conflicts with other duties or not.) The grounds determining whether one duty or obligation overrides another, then, lie partly in facts about *A*'s *doing x*, not solely in facts about the kind of action doing *x* is, independently of who does it.

If *A* ought<sub>c</sub> to do his duty to do *x*, the reason is that his doing so contributes to the accomplishment of some task that someone ought<sub>c</sub> to do. What is important about the task is that it be done, whether by *A* or by someone else, but what overrides or is overridden is *A*'s *doing x*, not the impersonal *getting x done*. Let *x* and *y* then, be any duties that conflict and are components of the same task—say, *z*. If to get *z* done it is more essential that *A* do *x* than that he do *y*, the former overrides. If it is necessary for him to do both to get *z* done, then it is not the case that he ought<sub>c</sub> to do either; for here *z* has an unperformable necessary condition and *A*'s effort is idle. *A* ought<sub>c</sub> to do either—it makes no difference which—if either suffices for *z* to be done. (That is, he ought<sub>c</sub> to do *x* or to do *y*—not, of course, that he ought<sub>c</sub> to do *x* or he ought<sub>c</sub> to do *y*. Here perhaps arbitrary rules, which abound in the laws of contracts and the regulations governing stations, will prevent indecision.)

Even if duties are components of different tasks, somewhat the same principles are operative. If *A*'s doing *x* is essential to the one task, his doing *y* being incidental to the other, then again the former overrides. Similarly, if the different tasks to which the two contribute are components of a further one, the issue concerns their necessity and sufficiency for getting that further task done. A single criterion to resolve conflicts exists only if there is some one task that every human being ought<sub>c</sub>

<sup>12</sup> Promises to marry deserve mention in their own right. At one time English law held that a man's breaking such a promise was punishable as breach of contract; it is not now so held, and people generally count a change of heart as sufficient excuse.

According to the theory of contracts given above, that a man ought<sub>c</sub> to keep a promise to marry presupposes that a woman has a right to marry. If a change of heart now excuses breaking such promises, people do not ascribe that right to women. The theory seems to fit facts. In the 19th century, marriage constituted the normal and expected means of survival for middle class women—working not being socially acceptable. (The case is different for proletarian women, but English law is not formulated to protect them.) In such circumstances, the right to marry inheres in the right to live. When it becomes acceptable for women to support themselves independently, the right evaporates.

Certainly, promises to marry are not given less seriously now than in the 19th century, and the change in attitude does not arise from that cause. So also, the same conventions now govern promising to marry; only they have less force.

to do—this task being antecedent to the allotment of duties, not itself something anyone has a duty to do. All conflicts could be resolved by finding the relations of duties to its accomplishment.

The environment of actions being complex, the resolution of problems about overriding is messy; yet the principles needed are simple. Always what an agent must discover is the relation of his doing a duty to the accomplishment of some task. Such problems belong to casuistry. It is important here not to confuse the question, "Does this duty override that?" with the different question, "Ought<sub>c</sub> the task to which this duty contributes be done?" Knowing that one duty overrides another presupposes, but is not a criterion for, knowing that some task ought<sub>c</sub> to be done. Casuistry rests on the foundations of ethics and not *vice versa*.

To summarize the positive features of the relations of duties and obligations to "ought<sub>c</sub>": First, if someone ought<sub>c</sub> to do *x*, then that *A* has an overriding duty to do it determines that he ought<sub>c</sub>. Secondly, if doing *x* is a component of *z*, and someone ought<sub>c</sub> to do *z*, then *A* ought<sub>c</sub> to do an overriding duty to do *x*. The grounds determining that a duty is overriding lie in the relation of *A*'s action to *getting z done*.

## IX

It was noted earlier that the original sin of the *prima facie* view is to suppose that an agent determines by a mere ritual what he ought<sub>c</sub> to do. By noting what the position overlooks in the relation

of duties and obligations to "ought<sub>c</sub>," it is possible to account for the temptation to sin.

The view overlooks the connection of duties with the division of labor and of obligations with the exchange of private property. By so doing, it by-passes the need *to show* that someone ought<sub>c</sub> to do the task to which doing one's duty contributes and that someone has a right to the kind of exchange which keeping one's obligation helps maintain. It assumes the acceptability, in the absence of countervailing considerations, of the framework in which duties and obligations are lodged. Private property and a division of labor being part of the way of life of a society, things absorbed with mother's milk, it is easy to omit the consideration of them. Generally people do not think about their way of life but merely act as others about them do.

It was Hume who took it as a principle of inquiry in morals that one should not get to the bottom of things, asserting that there is a rationality only of means and not of ends. Hume's principle is to dissolve the foundations of ethics in casuistry. There is a creeping Humeanism in all ordinary thought about what one ought<sub>c</sub> to do. The "ought" claims about which people principally deliberate—deliberating about others because of their relation to these—are of the sort asserting that a definite agent ought<sub>c</sub> (or ought<sub>c</sub> not) to do a definite action. Now it requires casuistry to determine the truth of such claims; there is a temptation to imagine that casuistry suffices. Actions fettered by worries about what general tasks people ought<sub>c</sub> to do mire down. The *prima facie* position is simply an effort to legitimate ordinary unconcern about the foundations of ethics.

*Eastern Washington State College*

*Received April 25, 1973*

## VII. UNIVERSALIZABILITY AND JUDGMENTS OF TASTE

JOHN FISHER

ONE of the constantly recurring assumptions in philosophical aesthetics is the hunch that, as Edmund Burke put it, the standard of taste is like the standard of reason, the same in all human creatures. Whenever this sort of assumption is operative in even the most peripheral way, philosophers explore the possibility of developing some vigorous structure for making judgments, or, in more impressive language, explore the logic of taste. There are many puzzles associated with the logic of taste, only one of which this paper will attempt to unravel.<sup>1</sup>

It is, I think, part of the heritage of the era of positivistic philosophy that taste appears so seldom as a serious notion or a significant term in contemporary aesthetics. To be sure the word occurs from time to time, but almost without exception in a sense which acknowledges the reductive effect of sceptical thought. It appears centrally in Frank Sibley's *Aesthetic Concepts*, but clearly only in his sense, that is, as an ability to see and tell something about an object. In this sense it is only negatively governed by conditions, so that I can say "*x* is garish implies that it is not the case that all of its colors are pastels," but otherwise shows no relationship at all to any conditions whatsoever. This sense has no attachment to any of the historically rich, albeit often mistaken uses of taste which are revealed by even the most superficial examination of 17th, 18th, and 19th-century aesthetic writings.

Now that is not necessarily bad. Perhaps everything has been said about taste and whatever relevance it may have to the philosophy of art. If such is the case this paper and whatever thinking it may provoke will nail the coffin lid down more securely. If reflection indicates the contrary, however, then the obituary was premature.

An initial look at the way "taste" has been used will show that, happily or unhappily, it is a most ambiguous term. Thanks to the elasticity not only of English but perhaps of all European

languages, it can imply both criticism and perception, standards and enthusiasm, subjective sensitivities, and objective order, and can apply both to those who know not only what they like but also why they like it, and those who know only what the masses in their culture like or choose.

There is nothing pejorative about labeling a word as multi-meaningful so long as the meanings can be sorted and distinguished (then it becomes "rich" not "ambiguous"). The test is a simple one: Do the meanings lead to disputes? A look at "taste" clearly shows that it has not been a rich term but a muddled one, giving rise to numerous quarrels, especially in the discussion about intuitive feeling versus rational judgment. Most American philosophers return to Hume and Kant when they think of questions of taste, passing over the heady controversies about feelings and reasons in Italy and France which preceded them.

Hume's *Essay on the Standard of Taste* admitted to the diversity of tastes and the disputes about taste, yet also to the existence of general principles, that is, general propositions about qualities of objects which yield pleasure to a qualified percipient. The explanation was given in Hume's denial of beauty as a quality of objects, insisting that it was a property of the contemplating mind. No sentiment represents what is really in the object. Each mind perceives a different beauty. Yet, Hume asks, do we not all recognize certain obvious differences, for instance, that Addison is better than Bunyan? Of course. For although the principles of taste are universal, the organs of internal sensation are seldom perfect enough to allow a feeling correspondent to those principles. Thomas Reid, when critical of such excessively subjective views, argued in this vein: If, in saying something is beautiful, I am not saying anything about it, but of myself, that is, my feelings, am I not confusing everyone? Why not simply say I like it?

There is something so dreadfully boring about

<sup>1</sup> The author would like to thank Monroe C. Beardsley, Alan Tormey, and D. Paul Snyder for their helpful comments on this essay.

that old saw, *Gustibus non est disputandum*, that one might do well to respond (in a moribund language of course), *Nihil praeter gustum est disputandum*, or perhaps better, *De gustibus solis est disputandum*. After all, you do not dispute facts, do you? Of course every student in an introductory philosophy class has learned how to untangle that one: sharpen up your expression, analyze the concepts of taste and what it is to dispute something. When you have done that you have no more puzzles. For those beyond the introductory level, however, let us turn to Kant.

"In order to distinguish whether anything is beautiful or not," said Kant at the beginning of his third Critique, "we refer the representation not by the understanding to the object for cognition, but by the imagination to the subject and its feeling of pleasure or pain." The judgment of taste (and taste for Kant is the faculty of judging the beautiful) then is not a judgment of cognition and is consequently not logical but aesthetic, that is, its determining ground is exclusively subjective. It does not refer a representation to an object as a property but to a subject and its feelings, its satisfaction. Kant knows that what appear as aesthetic judgments are usually just reports about pleasure felt; but genuine judgments of taste do exist and differ from other judgments in many respects, most significantly in "interest." Interest is a satisfaction related to desire, and it is central to Kant's intention to universalize the judgments that taste be seen as the faculty of judging an object, or representing it, by a totally disinterested satisfaction.

But judgments of taste are yet judgments, and all of the categories of analysis relating to judgments are applicable. Not only do we distinguish them from cognitive judgments in terms of the quality of disinterestedness but we must consider quantity, modality, and purpose as well.

Within these categories the uniqueness of Kant's analysis and some of its most troubling logical puzzles develop, for Kant must argue that the beautiful pleases universally without requiring a concept, as cognitive judgments demand. The judgment is subjective but is not a report of individual preference. The disinterested satisfaction implies a ground of satisfaction for all men. A judgment of taste about an object must claim validity for all men without this universality being rooted in the object.

Of course this is not to speak of pleasure, which is private and not corrigible. I like Bach, you like

Rock. Not all gentlemen prefer blondes. *Chacun à son goût*, the taste of sense. Beauty is quite a different matter. If beauty means anything at all it supposes in others comparable satisfaction. "*x* is pleasant to me" makes perfectly good sense. "*x* is beautiful to me" is, if not absurd, at least silly. I think Kant is perfectly correct at this point, and emotivists should perhaps avoid the term beauty entirely because it becomes systematically misleading in emotivism, just as *truth* is a muddling word in certain kinds of pragmatism. Perhaps I am here more hard-nosed than Kant, for he does not object to using "He has good taste" as suggesting that one knows how to please his guests. The rules which emerge here, however, are not universal but merely general, like empirical ones, for indeed they *are* empirical.

Think for a moment about the perfectly sensible sentence, "That is in bad taste." Now every instance I can think of suggests something about time and place. Jokes about morticians or corpses at a funeral, wrong time and place; bad taste. Obscenities are by definition in bad taste— notions which belong to the bedroom or the gynecologist's office; bad taste.

Good taste? Do we not simply mean the right amount of the right thing at the right time? Kant simply acknowledges usage here, but again, the rules are general. They are empirical, the universalization problem remains. Good or bad taste is inescapably relative.

But the interesting issues about taste are not the problems of good taste or bad taste but of the judgments of taste, such as "*x* is beautiful." Kant is understandably concerned about the status of such remarks. He is certain that they are subjective, but equally certain that they are uttered with the implicit conviction that others ought to agree, that somehow a principle of universalizability is entailed in aesthetic judgments. But this ought is not the ought of moral or cognitive judgments. Judgments of the good, for instance, make rightful claims to validity universally, but the good is represented by means of a concept as the object of a universal satisfaction. This is not the case with a judgment of taste. There is no rule under which anyone would be forced to call anything beautiful, no aesthetic analogue to the categorical imperative. Nevertheless, in a judgment of taste what is postulated is a universal something—Kant calls it a *voice*. It is not in fact the agreement of anyone. It only imputes this agreement. Not that people do agree about taste. Empirically

this is false: they do not agree. Yet a certain necessity prevails; not a theoretical objective necessity (for then everyone would feel the satisfaction in the object which I call beautiful): nor a practical necessity (which would imply some uniform action response), but an exemplary necessity, a necessity of the assent. But this cannot be standard apodictic necessity, and certainly it is not gained from experience. Required is a notion of commonality, not traditional common sense which judges by concepts, but one which judges by feeling. Kant's argument for the common sense seems to be circular, but I do not intend to discuss that point here. Germane to our concern is that this common sense (not external) is best thought of in terms of the effect of the free play of our cognitive powers. Look at the object and you will see that it purports to be beautiful for all because a beautiful object is an object whose form causes an awareness of the harmony of free ranging mental powers. The *sensus communis* requires that everyone ought to agree.

In sum, then, a judgment of taste is the judging of an object by a disinterested satisfaction which we call beautiful, something which pleases universally without being placed under a concept, whose form is the purposiveness of the object without a representation of purpose, and which is cognized as the object of necessary satisfaction, mind taking pleasure in the harmony of cognitive faculties. The paradoxical "purposiveness without purpose" aspect of taste judgments seems to take the object into account once more. Whenever an object pleases me aesthetically it looks as if it were designed to meet certain of my human needs, but of course this cannot be rationally defended. It appears as if it were purposively related to my desires, but the *as if* indicates that only the happy interplay of my faculties is involved. If this is so the object is quite irrelevant. We experience beauty then, not as a quality of the object, for beauty apart from any reference to the feeling of the subject is a cipher, but *as if* it were a quality of the object. In that sense the object is relevant after all.

Now if this sketch of Kant's treatment is at all accurate, some difficulties must be rising to the surface of all that germanic rumination. One of these, the old *gustibus* problem, Kant faces with cold confidence. There is an antinomy of taste to add to the collection of the *Critiques*. Everyone's taste is subjective, private, his own, and surely there is no disputing about taste. Yet we do quarrel about taste, and wherever we permit quarrels we

allow for the possibility of adjudication, and then the grounds of our judgment cannot be purely subjective. Well, that is an antinomy: judgments of taste are not based on concepts, otherwise they would be open to proofs. But judgments of taste are based on concepts, for otherwise we could not quarrel about them.

Now it seems to me that a number of alternative escapes from the antinomy are available, and Kant chooses a solution that is consistent with his transcendental philosophy but (or some might say, therefore) is more obscure than we would want. Stated simply, the judgment is based on a concept (the concept of the general ground of the subjective purposiveness of motive for the judgment) from which we can know nothing. Because of this the judgment is universalizable only because its ground lies in something like the supersensible substrate of humanity, ontologically transcendent, and inescapably Kantian. In simplest terms, the thesis should read, the judgment of taste is not based upon determinate concepts, and the antithesis should read, the judgment of taste is based on an indeterminate concept, the supersensible substrate, and that does not contradict the thesis. Now you must buy the whole Kantian philosophy for that escape. It is by no means clear that there is any obligation to avoid the antinomy in just that way, however. The problem is one for those who deny as well as those who accept the Kantian metaphysic.

What occasioned all of this for Kant is the perennial philosophical problem of universalization. The judgment of taste (*x* is beautiful) claims to be necessarily valid for everyone, but it cannot be shown to follow from a concept. Kant slips out of the dilemma by allowing the judgment to *refer* to a concept but denying that it can be *proved* by a concept. The concepts of the understanding are determinable through predicates of sense intuition, but the concept of the supersensible is transcendental and undeterminable.

Now let us step back and try to see what does follow from judgments of taste, particularly looking at that problem of universalizing judgments. Those who, with perhaps good reasons, at this stage find the heady prospect of universalizing any judgments bizarre are urged not to desert these considerations before certain distinctions are made. Some of the Kantian insights will be preserved, but I do not intend to lean on his ontology.

We have seen that the ambiguity or density or

richness, as you will, of taste is more complicated than one might at first glance suppose. The question is not merely, how does it operate, that is, how do we make judgments of taste (by an inner sense, by application of rules, learned or innate, etc.) but what is the judgment itself? Let us examine several quite different notions of what it is in which the judgment consists, before coming to the issues of implication and entailment.

(a) A judgment of taste is a liking or, perhaps better, reports a liking. Now likings are always interested. We like things which please, which satisfy, which we enjoy. This is totally and irrevocably subjective. *a* makes a taste judgment about *b* implies nothing in the way of obligation on anyone's part.

(b) A judgment of taste is a preference judgment, a *ranking* of alternatives. The criteria which are applicable are either subjective or objective. When interested they are subjective. Why do I rank Mozart's G Minor Quintet, K 516 higher than Beethoven's G Major Quintet no. 29? Subjectively I might say, with some truth, that it moves me more deeply. Objectively I might point out formal superiority with no reference to feeling at all, or I might show what objective features consistently make me feel delight.

But not all judgments of preference are interested. One can prefer an object for the qualities it possesses, apart from feeling responses. The evidence for this is straightforward: one can dislike his own tastes in art. No judgments need to correspond one to one with feelings. Rationalistic views of taste have insisted on this distinction between what I reason to be so about an object and the feeling response I experience in its presence. This objective aspect leads us to (c).

(c) A judgment of taste is a valuing, that is, a grading. A grading of objects, not of responses. It can be easily shown that I can value an art object aesthetically yet dislike it personally. This valuing is necessarily disinterested.

(d) A judgment of taste is a choosing, or at least provides something needed for an agent to make a rational choice. I may have a taste for Kolwitz and find Miro not to my taste. What might result is that I buy a Kolwitz and not a Miro when both are available. I actually choose. But I need not. I may not have the opportunity or the finances to choose. But I would so choose if I could. The choice is potential, just as I would

choose health over illness if I could, but there are inputs here completely out of my control. One of the disappointments of von Wright's admirable work *The Logic of Preference* is his reluctance to allow a fuller investigation of potential choice as an entailment of preference.

(e) A judgment of taste is a verdict, the conferring of value upon an object, a mental or linguistic act which makes an object beautiful by edict. Readers of contemporary aesthetics will recognize several cases immediately.

The status of a taste judgment does have something to do with the possibility of its being universalized. The question of what validly follows from a judgment of taste with respect to other persons is easily dismissed if a judgment of taste is exclusively an expression of liking, interested preference, or any other subjective notion. Unless I want to weaken the subjectivism by insisting like Hume that we are all somewhat alike (which is of course true, making hard-line subjectivism silly), my making a taste judgment about an object has no direct relevance to your judgments at all. The other cases are more interesting.

*a* makes a taste judgment with respect to object *b*. If you require a predicate, like Kant, call it beautiful. (I shall more often retain the broader and weaker expression, a taste judgment.) What follows if I reject the totally subjective premiss? Consider a parallel in moral philosophy: *a* makes a moral judgment (you may want to say he calls *c* good). What is he doing? Contemporary writers on moral theory have routinely standardized the alternatives:

- (1) He may be attempting to impart information about *c*.
- (2) He may be just confessing his likes.
- (3) He may be attempting to get to other people's feelings and change their behavior, as in emotivism. He likes *c* and wants you to like it too.
- (4) He might be expressing a moral principle, i.e., truth telling is good. He may be stating a special kind of imperative: Tell the truth, or Do your duty, etc.

Now the last of these reintroduces the universalizability problem. I happen to agree with those who hold that moral principles have some characteristics common with commands, not the least of which is that they must apply to other persons and situations which resemble the situation being refer-



red to. Not all commands are universalizable, but moral judgments are, for indeed that is their distinguishing feature. Not that people do what they are obliged to do. In fact they do not. I ought to do *c* does not imply I will do *c*. Nowhere is the distinction between descriptive and deontic accounts of moral judgments more apparent than at this point, and nowhere is it more important to distinguish between modalities which have to do with the truths of propositions (necessary and possible) and modalities which have to do with obligations (obligatory and permitted). Whatever logic of moral utterance one develops, one cannot allow obligatory *p* to imply *p* (whatever ought to be, is) nor the particularly insidious moral counterpart, *p* implies that *p* is permissible (whatever is, is all right). These have parallels in aesthetics which are less significant only to the extent that one's aesthetics is less important than his morals.

As a matter of fact the alternative answers to the question: What does one do in making a moral judgment? have their counterparts in taste judgments as well. Abbreviate "*a* makes a taste judgment about *b*" as  $T_{ab}$ .

- (1)  $T_{ab}$ : He may be trying to tell something about *b*. Objectivists and Formalists make this case.
- (2)  $T_{ab}$ : *a* may simply be confessing his likes. No reasons, therefore no disputes. *De gustibus non est disputandum*.
- (3)  $T_{ab}$ : He may be attempting to affect others' feelings and behavior. This is interesting, although the reasons may vary from the motivation of the critic to that of the parent trying to get the child to change the channel or to reduce the noise.

Note that none of these interpretations imposes any direct obligation on anyone. The question of generalization enters again at

- (4)  $T_{ab}$ : *a* might be implicitly making an ought statement. *a* makes a certain taste judgment about *b* implies that others ought to make the same judgment. In rather standard symbolic form,  $T_{ab} \rightarrow (\forall x)O_{xb}$ .

Like the moral argument, this cannot be a simple imperative. Unlike the moral argument, this cannot be an imperative at all. "Do your duty!" Very well, I can act on cue. "Like *b*!" Impossible. I cannot like on cue. "Judge beautiful!" Impossible. I cannot judge beautiful on cue. "Value positively!" I cannot on cue. (The distinction

between judging and uttering a judgment should be clear at this point. I can say "*b* is beautiful" upon command. I cannot judge *b* to be beautiful when ordered to do so.)

Consider this  $T_{ab} \rightarrow (\forall x)O_{xb}$ . (If *a* makes a taste judgment with respect to *b*, then for any individual *x*, *x* ought to make the same judgment.) Surely this move, to incorporate the modality into the predicate relation is logically suspect. The only possible justification for it would be to deny that this *ought* is the same ought as in moral obligation or the modal "obligatory." Consider the conditions of the implication. Materially the justification could be either (i) the correct aesthetic observer: the authority of *a*, he *knows* what ought to be a positive taste judgment, or (ii) the perfect aesthetic object: the optimum object, that *b* is what possesses the quality required. I doubt that many would want to accept one or the other disjuncts offered. In fact, they both have a strong scent of circularity about them. Despite the fact that they will not work, however, these considerations do bring to light the putative notion that "ought" is legitimately used in a number of different ways. There is a similarity of function but a distinct variability of substantive norms. (*a* is legally obliged to register for military service but may be morally obliged not to do so, etc.) Consider some instances of the obligation to refrain from certain acts:

Legal: You ought not park next to a fire hydrant.

If you do, you will violate section viii of the traffic code.

Moral: You ought not murder. If you do, you will violate norm viii of the moral code.

Rational: You ought not say both *p* and *not-p* at the same time. If you do you will violate axiom viii of the logical set.

Aesthetic: You ought not judge Correggio's *Jupiter and Antiope* to be bad. If you do you will violate aesthetic norm viii.

I do not suggest for a moment that the latter use of ought has any of the practical authority of the others, nor indeed that it can be hardened into a law-like principle without closing one's eyes to the obvious distinctive features of any talk about art (the problem of dealing with appearances, the so-called gratuitousness of works of art, the ease of overriding aesthetic judgments, etc.). In fact, the point to be made is precisely that this use of ought in aesthetics is different in many ways from other uses, yet it operates in much the same manner.

I make no logical distinction between "ought" and "should" here, and those who find the suggestion of "ought" in aesthetics discomforting can replace it with "should" with no consequent alteration of meaning.

It is not difficult to describe the norm of legal, moral, or modal obligation. We can point to each of them, citing volume and page in basic texts of legal, moral, or logical laws. (Of course, this is not equivalent to saying there is consistency among or universal acceptance of the norms.) The norms of the last category are much more difficult to pin down, but formally they function in the same way. The first three are, in practice, stated and meant to be understood in universal terms. No one (not even policemen) should park next to fire hydrants. No one should murder. No one should contradict himself. One is in each case committed to the application of the should beyond the individual I. In spite of the fact that judgments of taste are much more easily defeasible, they too entail the commitment to apply the judgment beyond the subjective I.

But what can this taste judgment have to do with another person? What can my judging a performance of a Beethoven string quartet to be beautiful have to do with you? One has only to remember Santayana's succinct objection to obligation at this point, namely that it is meaningless to say that one ought to see what one cannot see. Furthermore it could not suffice simply to state  $T_{ab} \rightarrow (\forall x)O_{xb}$ , not only because of the logical objections to the inclusion of any modality in the content or substance of the predicate, but because when I make a universal judgment I accept responsibility for giving some sort of reason for the judgment's being universalized.

Too, often, at this stage of investigation, we are distracted into a consideration of what counts as a good reason for saying that something is beautiful. As interesting and important as that might be, it is not the issue at hand. The issue is, what does my making a taste judgment commit me to? Commit here is used in a strong sense; judging  $T$  means that another judgment is obligatory, another judgment about other persons. Very simply, when I judge  $b$  beautiful I commit myself to the notion that if you could judge (not simply call)  $b$  beautiful, you would in some significant sense be better off, that is, some benefit would accrue to you as it has to me. When we make taste judgments we impute this being benefited to all those who make, or would make, like judgments and impute

some taint or blame to those who disagree, for they miss out on the benefits.

Making a judgment of taste is rarely a mere academic exercise. Even the professional critics, those perhaps most tempted by the pressures of their work to make at times judgments unrelated to the concept of benefits, almost never can be discovered without the explicit or implicit indication that the realization of the values which they ascribe to a work involves benefits. Nor is this to deny the possibility of intrinsic value to works of art. It has nothing to do with intrinsic value. In a universe in which all persons were blind one could say (indeed I would want to say) of a Caravaggio or a Cézanne: If anyone could see it, he would be benefited in some way.

Formally then I can adapt an entailment operator to capture the sense of "commits one to," so that making a taste judgment commits one to the notion that for any individual, if he makes a like judgment, he will be benefited in some way:  $T_{ab} \hookrightarrow (\forall x)(T_{xb} \rightarrow B_x)$ , where  $\hookrightarrow$  is read "commits one to." It does not strictly follow that my making a taste judgment will actually result in benefit accruing to anyone else, nor that everyone or even anyone will ever make a like judgment. What does follow is that in making a taste judgment I commit myself to a universalized conditional judgment. Circumstantial evidence for this conclusion is overwhelming. Consider the replies one hears to matters of disagreement in taste judgment. "That music is lovely." "I don't think it's at all lovely." What responses do we most frequently hear? "Well, that's your loss," or "If you'd stop listening to all that trash and listen to something good for a change, you'd think it beautiful too!" or "You don't think so? Well then let me explain it and then listen again and you'll see," etc. We can, and do say, "Oh well, each to his own taste" in response to disagreement, but even that is often said with a wistfulness which seems to border on pity, and indicates that the speaker believes that some benefit is lost by anyone who fails to make the same judgment.

Of course, the limits of what follows from this observation are severe and one must not be trapped in some familiar *non-sequiturs*. There is a tendency to move from  $(\forall x)(T_{xb} \rightarrow B_x)$  to a kind of preachment, "If you want to be bettered in some way, make these judgments." This would destroy even the weak sense of aesthetic ought, for this is no obligation at all. Furthermore there is the unwelcome possibility of insinuating a beneficial

ideal into the framework itself in the worst possible way. (After all, do we not believe that others would be better off if their judgments generally agreed with ours? Does not the ideal include this kind of agreement?) This would reduce the whole matter to question begging. Being benefited in the sense I want to use has nothing to do with an ideal. It has simply to do with human needs and interests in the broadest sense. Let the specific benefits be developed in whatever direction experience dictates. Let the psychologizing aestheticians define *B*. In any case the above seems to be the most reasonable candidate for what  $T_{ab}$  can fully mean. The common criticisms of aesthetic utilitarianism are not applicable, for this does not speak to the question of what use or benefit the work of art might have. Wilde could be right; all art could be utterly useless and this notion still be correct, for it speaks only of judgments, not of objects. The obvious prime candidate for the benefit, if one counts the number of pages devoted to it in the works in aesthetics, would be the stimulation or creation or occasioning of aesthetic experience, whatever that might be, but that issue is not my present concern. Suffice it to say that universalizability does obtain, and the predicate explains the aesthetic, that is, the practical sense of ought.

This has been an attempt to explore certain formal problems of aesthetic judgment. Not what makes something beautiful, not what makes me say it is beautiful has been my concern but, rather, what is going on when we make a judgment.

Nothing in this account suggests that  $T_{ab} \leftrightarrow (\forall x)(T_{xb} \rightarrow B_x)$  is necessarily and overtly characteristic of those who judge, nor that everyone who makes a taste judgment *acknowledges* that in so doing he commits himself to this universalized consequence. Rather, this commitment is made as a matter of fact, and, if the principle that one ought to make judgments which have benefits is a reasonable one, reasonable men are justified in making such a universal commitment. This allows, with conceptual frugality, something of what Kant wanted to obtain, and explains what he could handle only by means of a larger ontological commitment. Yet my conclusions are, surprisingly, not as far from his as one might have suspected. Kant demanded that a certain universalizability operation be allowed. His is mistaken in fact and has to be justified by an ontology which is ponderous and which demands a metaphysical commitment outside of aesthetics. I have suggested a simpler although no less "unproved" basis for the generalization, agreeing that taste judgments have to be more than simple preferences. If the route to a partial agreement with Kant was circuitous, the effort, like the journey to Ithaca, might have been the totality of the reward, although, as Mr. Weller, somewhere in the *Pickwick Papers*, wonders,

... whether it's worth goin' through so much to learn so little, as the charity-boy said when he got to the end of the alphabet, that is a matter of taste!

Temple University

Received July 2, 1973

## VIII. CLASSES, COLLECTIONS, AND INDIVIDUALS

RUTH BARCAN MARCUS

THE logical theory of sets obscures commonplace distinctions which we make between assortments, collections, and classes; in particular, those commonplace uses where we speak of collections of physical objects. In this paper, I shall suggest that the puzzles generated by assimilating ordinary collections to classes reflect the same difficulties as those which follow from equating proper names with definite descriptions. An indication will also be given as to how, within an alternative logical framework, these distinctions may be articulated.

Consider the following example from an elementary mathematics text.

In a club with five members, there are 5 possible committees with one member, 10 committees with two members, 10 committees with three members, 5 committees with four members, 1 committee with five members. The selection of a committee which includes all five members (sometimes referred to as the "committee of the whole") means that there are zero members not serving. Thus we may balance our table by saying that there is 1 possible committee with zero members. (*School Mathematics Study Group*, Yale University Press, 1960, Vol. II, Pt. II, p. 284.)

To take the above quotation seriously is to be puzzled. Clubs as we know them have constitutions and by-laws. It usually defines each of its committees, not by naming particular individuals as members, but by specifying its number and function, and how its members are to be selected (i.e., "The Executive Committee shall consist of the officers and one person elected by a majority vote of the membership. Its duties shall be . . .") Suppose  $a, b, c, d, e$  are the five members of the club described, and suppose the same three persons ( $a, b, c$ ) were elected to the Nominating Committee and the Resolutions Committee. According to the text quoted, they are the same committee. But are they? Would the Resolutions Committee as Resolutions Committee propose nominations? Would the Nominating Committee as Nominations Committee propose resolutions? Isn't it merely an historical contingency that the same persons serve on both just as it is an historical contingency

that  $a, b, c$  compose the Resolutions Committee? It might have been  $c, d, e$ . Similarly for the Nominating Committee. Yet, it is certainly true given the accomplished fact, that every member of the one is a member of the other. There is then a kind of accidental equivalence between those committees. But surely it is not an historical accident that the collection of persons  $a, b, c$  is the same as the collection of persons  $a, b, c$ . Indeed if the by-laws were such that each of these two committees was a committee of one rather than three, to which the same member had been elected, we would discern shades of the evening star and the morning star. The committees of the example correspond to classes where composition is fixed by describing conditions of membership. But the individuals  $a, b, c$  comprise a collection or aggregate of specific individuals. The theory of classes (or sets), by virtue of some of its fundamental assumptions, conceals such differences. Yet, "class," and "collection," in logic are not held to be mere homonyms of colloquial counterparts. In naïve and often in axiomatic set theory, they are introduced in terms of correlative informal notions. Departures from familiar meanings are taken to be technical extensions, modifications, or refinements. Consider the following examples:

- (1) The essential point of Cantor's concept is that a collection of objects is to be regarded as a single entity (to be conceived of as a whole). The transfer of attention from individual objects to collections of individual objects is evidenced by the presence in our language of such words as 'bunch,' 'covey,' 'pride,' and 'flock'. With regard to the objects which may be allowed in a set, (Cantor's description) "objects of our intuition or of our intellect" gives considerable freedom. . . . Green apples, grains of sand or prime numbers are admissible as sets. (R. R. Stoll, *Introduction to Set Theory and Its Logic*, pp. 2-3.)
- (2) A pack of wolves, a bunch of grapes, or a flock of pigeons are all examples of sets of things. . . . To avoid terminological monotony, we shall sometimes say *collection* instead of *set*. The word 'class' is sometimes used in this context. (P. Halmos, *Naïve Set Theory*, N.Y., 1960, p. 1.)

- (3) By a *set* we mean any kind of collection of entities of any sort. . . . Many other words are used synonymously with 'set': for instance, 'class,' 'collection,' and 'aggregate.' (P. Suppes, *Introduction to Logic*, N.Y., 1959, p. 177.)

The common definition of "collection" as

- (4) A group of things which have been gathered together.

seems adequate to familiar examples such as coin or art collections, or the sundry objects in a woman's purse. Yet there seems to be an inescapable peculiarity in the shift [as in (1)] to entities such as numbers as if they could be gathered together like pebbles on a beach. Is it merely an extension of a familiar notion to include objects of our intellect as members of collections in accordance with Cantor's proposal? Medals may be collected but what of virtues? The response to such naïve questioning is that no such literal gathering together is intended. As Quine puts it:

- (5) We can say that a class is any aggregate, any collection, any combination of objects of any sort; if this helps well and good. But even this will be less help than hindrance unless we keep clearly in mind that the aggregating or collecting or combining here is to connote no actual displacement of objects, and further that the aggregation of say seven given pairs of shoes is not to be identified with the aggregation or collection or combination of those fourteen shoes nor with that of the twenty eight soles and uppers. In short, a class may be thought of as an aggregate or collection or combination of objects just so long as 'aggregate' or 'collection' or 'combination' is understood strictly in the sense of 'class'. (W. V. Quine, *Set Theory and Its Logic* [Cambridge, 1963].)

But to remain unabashed by the circularity is not to resolve the perplexity. If abstract objects such as numbers are eligible to be elements of collections, does it make sense to talk of displacement at all? And what principle precludes the identification of an aggregate of seven pairs of shoes with those 14 shoes? It would be difficult to persuade a shopkeeper.

Implicit in the ordinary definition of "collections" (4) is that (a) collections are not empty and (b) they consist of distinct and distinguishable objects. Also, since "gathered together," "aggregated" are in the past tense, (c) collections are in some sense complete. But what kinds of objects, how are they distinguished, and in what sense

have they been "gathered together"? Confusion is minimized if we restrict the entities of a collection to individuals which are distinguishable as physical objects. And indeed this is the colloquial use of the term. Having been gathered together, the number of objects is finite, whether they be assembled in heaps, piles, or hung on museum walls. Nevertheless, even here "gathered together" is difficult to make precise. Generally spatio-temporal proximity is involved though, as in the case of an art collection, not necessarily. The common feature of such aggregates is that they consist of enumerably many individual physical objects. How many? Presumably some finite plurality depending on the particular mode of aggregation. How many grains of sand make a heap, or grapes a bunch?

We do not have a common word for

- (6) An *arbitrary* finite selection of physical objects •

which is entirely separable from connotations about the *mode* of aggregation. In the absence of a more felicitous alternative, I shall use "assortment" for (6). I do not propose here to go into an analysis of the notion of physical object or to raise questions as to whether, for example, shadows or after-images qualify. I shall assume that it is a sufficiently well understood predicate so that particular physical objects can be individuated.

It is clear from the above that logical classes are very different from assortments. They may consist of any entities whatever including entities which may themselves be classes. Logical classes may be empty or infinite. The number of members of a logical class, even where the members are taken to be physical objects, may be indefinite in an historical sense as when we speak of the class of animals or stars, since such entities are still being born.

An exhaustive account is given of an assortment by specifying each of its members. But the extended notion of class precludes such an account in almost all cases. If the common definition of collection as in (4) is to apply to logical classes as well, it must be in some other sense of "gather" as when Kant speaks of "gathering together under a concept" or "gathering together under a rule." The theory of classes or sets obscures the distinction between these two ways of "gathering together," or, using Cantor's phrase, these two modes of conceiving a totality.

What I have suggested above is that there is an

ambiguity in the notion of "gathering together" and "conceiving a totality" which is reflected in the distinction between an itemization of elements and a statement of conditions for membership. The obliteration of the difference by the logic of set theory does not generate too many perplexities for mathematics (leaving aside for the moment paradoxes of the Russell and Skolem sort) since the elements with which mathematics deals are *defined* by those conditions. A number cannot fail to satisfy any one of its unique conditions and still be that number. It makes no sense to say that 3 is no longer the whole number between 2 and 4 but is still 3. But it does make sense to say that Venus has changed its orbit but is still Venus. An entity such as a number may therefore be equated with any one of its unique conditions without generating many of the familiar puzzles which follow upon equating the proper name of an object such as a physical object with a definite description (other than a definite description which specifies as the unique condition that it is that object, e.g.,  $\lambda(x)Ia$  where '*I*' is for the identity relation between individuals and '*a*' the individual object's proper name).

The absence of differentiation between these two modes of conceiving a totality is apparent from the manner in which sets are described. For example, one reads in Suppes' *Introduction to Logic* (N.Y., 1959):

- (7) Often we shall describe a set by writing down the names of its members, separated by commas and enclosing the whole in braces. For instance by

(a) {Roosevelt, Parker}

we mean

- (b) the set consisting of the two major candidates in the 1904 American presidential election.

By

(c) {1, 3, 5}

we mean

- (d) The set consisting of the first three odd positive integers. (P. 179.)

In (7a, c) we find an exhaustive list of names which is bracketed, whereas in (7b, d) what is bracketed is a specification of conditions to be met. Yet (7a) is taken to have the same *meaning* as (7b), and similarly (7c) and (7d). No great difficulties are encountered in connection with the numerical case if we do not take a numeral to

have the referential function of a proper name. But here, too, if the numerals *are* taken to be proper names in a strict sense, and we attempt to track down the objects of reference, then antinomies develop.<sup>1</sup> It would be much like trying to find the object of reference of the White Queen of chess. Any *candidate* whatever for White Queenship must move in accordance with the rules, and beyond that there would be no way of deciding between rivals for the job.

The more familiar difficulties are connected with equating (7a) and (7b); taking as synonymous a designation of an assortment which, as suggested above, consists of an inventory, and a description which fixes a range of objects, not by an inventory but by specifying the conditions to be met by undesignated objects. A well known example will serve to remind us of these puzzles.

According to (7)

- (8) (a) {*N*, *M*, *V*, *P*, *MA*, *U*, *E*, *S*, *J*} (where '*N*', '*M*', ... abbreviate the proper names of the planets)

means the same as

- (b) the set of planets.

But surely not. Venus' membership in the assortment designated by (8a) is a matter of logical necessity, but not its membership in the class of planets. The number of objects in the assortment (8a) is *logically* fixed, not so the number of planets. We may know that Venus is contained in an assortment which has Venus among its elements, and not know that it is a planet and so on and so on.

But (7) should not be taken as a truth claim about synonymities which have been discovered to hold between expressions designating assortments by inventory and expressions which describe classes without specification of individual members (although it is sometimes mistaken as such a claim). Rather, proposals such as (7) should be understood as a decision to ignore the difference between what we have called assortments and logical classes. By virtue of such a decision, we arbitrarily restrict talk of classes to truth functional contexts. If we adhere to this decision, we would, for example, allowing for the possibility that by some cosmic cataclysm Venus might be hurled out of its orbit around the sun, have no way of asserting that it is possible that Venus not be a

<sup>1</sup> See for example, the interesting paper of Paul Benacerraf, "What Numbers Could Not Be," *The Philosophical Review*, vol. 74 (1965), pp. 47-73.

planet. Such an assertion would generate a contradiction. The systematic manifestation of principles such as (7) is, of course, the postulation of a principle of extensionality; that classes which have the same members are the same class. There is an obviousness about such a principle, but only because the intuitive notion of a class is what we have termed an assortment. And indeed, for assortments, extensionality need not be postulated at all. But logical classes are not conceived of as an actuality of finitely numerable objects. Rather, they are generated by a general principle of set formation. Such an *Aussonderungssaxiom* states roughly that

- (9) To every *wff*  $A$  which contains  $a$  as the only free variable, there corresponds a class  $r$  and to every class  $r$  there corresponds a *wff*  $A$  such that  $A = a \in r$ .

If the device of abstraction is used,  $r$  is the abstract of  $A$ . The symbol '=' may as in *Principia Mathematica* signify equality by definition in which case the abstraction operator ' $\wedge$ ' and the connective for membership ' $\in$ ' are taken as part of the contextual definition of classes. Where (9) is introduced axiomatically the abstraction operator is generally taken as primitive, and '=' is a systematic relation of equivalence such as material equality, or a stronger equivalence if it is available as in some of the modal systems.

It is well known that (9) is not acceptable unqualifiedly in view of Russell's paradoxes and the perplexities stemming from the Lowenheim-Skolem theorem. Our choice of a solution is type-theoretic, not merely because it resolves the Russellian paradoxes, but also because the notion of an assortment is grounded in the supposition that there are at least individual *physical* objects which are not themselves sets or classes.

Consider such a type-theoretic applied predicate calculus of second order with individual and predicate variables and constants, and abstraction. Individual constants are interpreted as proper names of physical objects. Under this interpretation, identity is a relationship which holds between an individual and itself. As such, an identity claim is not a contingent matter. We are supposing here that there is a purely referential function of expressions which is the primary

use of proper names.<sup>2</sup> A proper name does not have a predicative function whereas a singular description predicates uniquely. Although self-identity is not a contingent matter, it may be a contingent matter that two singular descriptions apply to the same individual. This suggests that there are weaker equivalence relations proper to descriptions. The systematic representation of these differences is roughly to construe singular descriptions as (1) unary abstracts on the level of first order predicates which (2) contain a uniqueness condition and (3) contain predicate expressions other than ' $I$ .' (The latter excludes for example " $\hat{x}(xIa)$ " as a singular description. Rather,  $\hat{x}(xIa)$  as we shall indicate below is a single-membered assortment.) The discussion of proper names is not merely tangential to the consideration of classes and assortments. For the analogies are close. If, to extend a metaphor of Searle's,<sup>3</sup> proper names are pegs on which to hang (singular) descriptions, then a bracketed list of proper names (which designate an assortment) are pegs on which to hang class descriptions as in the case of (8a, b). Just as "the immense pragmatic convenience of proper names in our language lies precisely in the fact that they enable us to refer publicly to objects without being forced to raise issues and come to agreement on what descriptive characteristics exactly constitute the identity of the object"<sup>4</sup> so may we, through a device like bracketed lists of proper names, refer publicly to groups of objects, without being committed to any one identifying descriptive characterization of the assortment other than the one which characterizes it as an assortment. As in the case of proper names, an equivalence relation between assortments is never contingent but an equivalence relation between classes may be.

If we are to be able to represent such distinctions between various equivalence relations such as identity of individuals, logical, material, and contingent equality of classes within the object language of the predicate calculus, we shall need in addition to the truth functional connectives a stronger implication.

In a previous paper<sup>5</sup> I discussed the formal details of such a system ( $S_4^2$ ) which is based on an extension to second order of  $S_4$ . In it we have metatheorems which correspond to the following:

<sup>2</sup> See for example J. R. Searle's concise paper, "Proper Names," *Mind*, vol. 67 (1958), pp. 166-173.

<sup>3</sup> *Ibid.*, p. 172.

<sup>4</sup> *Ibid.*, p. 172.

<sup>5</sup> "Classes and Attributes in Extended Modal Systems," *Acta Philosophica Fennica*, fasc. 16 (1963), pp. 122-135.

- (10) A principle of extensionality holds for assortments.
- (11) A principle of extensionality holds for classes which are strictly equal. For materially but not strictly equal classes, the principle of extensionality is restricted.

An elaboration of (10) and (11):

- (12) Two classes may be weakly equivalent (materially or contingently). If each of them is weakly equivalent to some assortment, the corresponding assortments are the same.

A word of elaboration about (10–12). It should be noted that I said “A principle of extensionality” rather than “the principle of extensionality.” The strongest claim would be that if classes have the same members then they are identical. In  $S_4^2$  the identity relation is not defined over classes but over individuals. However, the systematic manifestation of an extensionality principle is an unrestricted substitution rule. Such a substitution rule is provable for assortments and for classes which are *strictly* equal. Our interpretation of  $S_4^2$  would still exclude oblique epistemic contexts since although two classes may be logically equal, the alternative descriptions may not be inter-substituted in such contexts without going from truths to falsehoods.

Returning now to  $S_4^2$ , the abstraction operator ( $\hat{\phantom{x}}$ ) is taken as primitive and an *Aussonderungsaxiom* postulated. For unary abstracts with no free variables, it takes the form (omitting precautions about choice of variables)

- (13)  $(b \in \hat{aA}) \equiv B$  where  $a$  is the only individual variable occurring freely in  $A$ , and  $B$  results from replacing all free occurrences of  $a$  in  $A$  by  $b$ .

The notation  $r, s, t$  will be used for unary abstracts. They may be understood as designating logical classes.

Definitions of sum, product, complement are introduced in the usual way. The two implication relations ( $\supset, \ni$ ) and the two equivalence relations ( $\equiv, \equiv$ ) have their Boolean analogues: material and strict inclusion ( $\subset, \in$ ), and material and strict equality ( $=_m, =$ ). We shall have reason to refer to the latter definitions which may be given as

- (14)  $(r =_m s) =_{\text{df}} (\forall a)((a \in r) \equiv (a \in s))$   
 (15)  $(r = s) =_{\text{df}} (\forall a)((a \in r) \equiv (a \in s))$   
 (or  $=_{\text{df}} (r =_m s)$ )

The null and universal terms are defined as

- (16)  $V =_{\text{df}} \hat{a}(aIa)$  and  $V =_{\text{df}} -V$

An assortment may be represented as follows:

- (17)  $\hat{a}A$  designates an assortment if there is a  $B$  such that  $A$  may be transformed into  $B$  by elimination of defined constants, and by application of rules of association, distribution, and double negation where  $B$  is of the form

$$\hat{a}((alb_1) \vee (alb_2) \dots (alb_n))$$

‘ $I$ ’ names the identity relation, and  $b_1, b_2, \dots, b_n$  are individual constants.

It can be shown by an induction on the number of disjuncts that a principle of extensionality holds for assortments:

- (18) Let  $r$  designate an assortment as in (17) of individuals  $b_1, b_2, \dots, b_n$  ( $n \geq 1$ ) and  $s$  an assortment of individuals  $c_1, c_2, \dots, c_n$  ( $n \geq 1$ ). If  $r =_m s$  then  $r = s$ .

For it is also provable that

- (19) If  $r = s$  then ‘ $r$ ’ may replace one or more occurrences of ‘ $s$ ’.

It should be noted that unrestricted substitution of ‘ $r$ ’ for ‘ $s$ ’ is not available where

$$(r =_m s) \ \& \ \sim(r = s).$$

However,

- (20) If  $r =_m s$  and if ‘ $s$ ’ does not occur within the scope of a modal operator, then ‘ $r$ ’ may replace one or more occurrences of ‘ $s$ ’.

By virtue of (20), in the case of material equality, substitution is permitted only in truth functional contexts.

The more formal statement of (12) reads

- (21) If  $r_1 =_m r_2$ ,  $r_1 =_m s_1$ ,  $r_2 =_m s_2$  and if ‘ $s_1$ ’, ‘ $s_2$ ’ are abstracts which designate assortments, then  $s_1 = s_2$ .

Returning now to the original example. The Nominating Committee may be represented by an abstract over a set of defining conditions specified in the by-laws. Call it ‘ $N$ ’. Similarly, the Resolutions Committee ( $R$ ). The assortment (or collection or aggregate) consisting of the three members  $a, b, c$  may be represented as

$$(22) \hat{x}[(xIa) \vee (xIb) \vee (xIc)](C)$$

Then if

- (23) (a)  $(R =_m N) \ \& \ \Diamond(R =_m N)$   
 (b)  $(R =_m C) \ \& \ \Diamond \sim(R =_m C)$   
 (c)  $(N =_m C) \ \& \ \Diamond \sim(N =_m C)$



' $N$ ', ' $R$ ', ' $C$ ', are intersubstitutable in, for example

(d)  $(a \in R)$

but not in

(e)  $(R = R)$

for, when expanded, it is seen that in (23e) ' $R$ ' is in the scope of a modal operator (15). Suppose furthermore that  $a_1$ ,  $b_1$ ,  $c_1$  are alternative proper names of  $a$ ,  $b$ ,  $c$  respectively. Call the assortment ' $D$ '. Since

(f)  $(N =_m D)$

it follows from (22 a, b, f) that

(g)  $C =_m D$

But ' $C$ ' and ' $D$ ' designate assortments and from (20) it follows that

(h)  $C = D$ .

This is not a case of converting contingencies into necessities by symbolic hocus-pocus. That  $a$ ,  $b$ ,  $c$  by whatever name we call them happen to have been elected to two committees is interpretable as a contingent fact. But it is not a contingent matter that however we tag them, they are a single assortment. Indeed isn't this a more suitable analysis of the principle of extensionality; that it holds for assortments and perhaps collections but not classes. Classes are, in Frege's language, "unsaturated."<sup>6</sup>

*Northwestern University*

*Received September 3, 1965*

<sup>6</sup> This paper was accepted for publication in 1965. The retarded publication is due to the author's (ultimately negative) deliberations regarding revision. The journal is not responsible for the delay.

## BOOKS RECEIVED

- de Berredo Carneiro, Paulo E. and Arnaud, Pierre (eds.), *Auguste Comte: Correspondance et Confessions* (Mouton: Paris, 1973), 437 pp.
- Caton, Hiram, *The Origin of Subjectivity* (Yale University Press: London and New Haven, 1973), 248 pp., \$12.50.
- Corti, W. R. (ed.), *The Philosophy of George Herbert Mead* (Amriswiler Bücherei: Switzerland, 1973), 260 pp.
- Goetzmann, W. H. (ed.), *The American Hegelians* (Alfred A. Knopf: New York, 1973), 397 pp. \$15.00.
- Hooker, C. A. (ed.), *Contemporary Research in the Foundations and Philosophy of Quantum Theory* (D. Reidel Publishing Co.: Dordrecht, 1973), 385 pp.
- Hyman, Arthur and Walsh, J. J. (eds.), *Philosophy in the Middle Ages* (Hackett Publishing Co.: Indianapolis, 1973), 747 pp., \$5.95.
- Kenny, Anthony, *Wittgenstein* (Harvard University Press: Cambridge, Mass., 1973), 240 pp.
- Lorenz, Krüger, *Der Begriff des Empirismus* (Walter de Gruyter: Berlin & New York, 1973), 283 pp.
- Kurtz, Paul (ed.), *The Humanist Alternative: Some Definitions of Humanism* (Prometheus Books: Buffalo, 1973), 190 pp., \$8.95.
- Lackey, Douglas (ed.), *Essays in Analysis: Bertrand Russell* (George Braziller: New York, 1973), 345 pp., \$8.95.
- Leitner, Bernard, *The Architecture of Ludwig Wittgenstein* (The Press of the Nova Scotia College of Art and Design: Halifax, 1973), 127 pp., \$9.95.
- Narski, I. S., *Dialektischer Widerspruch und Erkenntnislogik* (VEB Deutscher Verlag der Wissenschaften: Berlin, 1973), 440 pp.
- Nørreklit, Lennart, *Concepts: The Nature and Significance for Metaphysics and Epistemology* (Odense University Press: Odense, 1973), 226 pp.
- Pearce, Glenn and Maynard, Patrick (eds.), *Conceptual Change* (D. Reidel Publishing Co.: Dordrecht, 1973), 282 pp.
- Petőfi, Janos S. and Franck, Dorothea (eds.), *Präsuppositionen in Philosophie und Linguistik* (Athenäum Verlag: Germany, 1973), 606 pp.
- Proceedings of The Aristotelian Society*, Vol. LXXIII (Methuen & Co.: London, 1972/73), 251 pp.
- Rescher, Nicholas, *Welfare* (University of Pittsburgh Press: Pittsburgh, 1972), 186 pp., \$7.95.
- , *The Coherence Theory of Truth* (Clarendon Press/Oxford University Press: Oxford, 1973), 374 pp., \$13.00.
- Rigobello, Armando, *Ricerche sul Trascendentale Kantiano* (Editrice Antenore: Padova, 1973), 277 pp.
- Scharfstein, Ben-Ami, *Mystical Experience* (Bobbs-Merrill: New York, 1973), 195 pp., \$6.95.
- Twersky, Isadore and Williams, G. H. (ed's), *Wolfson: Studies in the History of Philosophy and Religion*, Vol. I (Harvard University Press: Cambridge, 1973), 626 pp., \$15.95.

# METAPHILOSOPHY

edited by TERRELL WARD BYNUM and WILLIAM L. REESE  
assisted by JOHN SPISSINGER

## CONTENTS

|  |                     |
|--|---------------------|
| Normative ethics that are neither teleological nor deontological . . . . .                                 | DONALD R. KOEHN     |
| Neutrality . . . . .   | MICHAEL P. T. LEAHY |
| Foot and Hare on naturalism . . . . .  | MALCOLM B. E. SMITH |
| Ontological commitments of everyday language . . . . .   | CHARLES CRITTENDEN  |
| Particulars and ontological parity . . . . .   | DOUGLAS GREENLEE    |
| Back to reality . . . . .  | H. GENE BLOCKER     |
| <i>Review</i>  |                     |
| Thomas Nagel: <i>The possibility of altruism</i> . . . . .   | ROBERT AUDI         |
| David Lyons: In the interest of the governed: a study in Bentham's philosophy of utility and law . . . . . | ROBERT FULLINWIDER  |
| The Philosopher as Teacher   |                     |
| Logic and rhetoric . . . . .   | D. V. C. LINCICOME  |
| Comment: A Change in the Weather . . . . .   | ROBERT C. SOLOMON   |

Annual Subscription \$15.00 (Institutions); \$10.00 (Individuals)

Published for the METAPHILOSOPHY FOUNDATION INC. by BASIL BLACKWELL, Oxford

Orders to: Journals Department, BASIL BLACKWELL & MOTT LTD., 108 Cowley Road, Oxford OX4 1JF, England

## ALAN ROSS ANDERSON MEMORIAL FUND

Alan Anderson died last December 5th, of cancer, at the age of 48. As a memorial to his manifold contributions (both academic and personal), there has been established at the University of Pittsburgh an Alan Ross Anderson Memorial Fund whose initial purpose is to fund a prize or fellowship for a graduate student or junior faculty member in philosophy. Readers of this journal who wish to remember him in this way should send a check to the **Alan Ross Anderson Fund**, in care of the **Department of Philosophy, University of Pittsburgh, Pittsburgh, PA 15260.**

# RATIO

edited by STEPHAN KÖRNER

Volume XVI

June 1974

No. 1

GÜNTER SCHULTE  
OTA WEINBERGER  
VADIM SADOVSKY  
PETER ALEXANDER  
JOHN HARRIS  
ALBERT MENNE  
HILLEL STEINER  
PHILIP MULLOCK  
M. W. BEAL  
DIETER WANDSCHNEIDER  
I. L. HUMBERSTONE  
J. N. HATTIANGADI

Kant and the Problem of the Critique of Reason  
Contrary-to-Fact and Fact-Transcendent Conditionals  
Problems of a General Systems Theory as a Metatheory  
Boyle and Locke on Primary and Secondary Qualities  
Leibniz and Locke on Innate Ideas  
What is Truth?  
The Concept of Justice  
The Permissiveness of Powers  
Essentialism and Closed Concepts  
Concerning the Antinomy Problem of Logic  
Logic for Saints and Heroes  
The Importance of Auxiliary Hypotheses

Price £2.25 net

Annual Subscription £4.00 net

BASIL BLACKWELL  
OXFORD, ENGLAND

Vol. 4 No. 1 April 1974

## Journal of the Theory of Social Behaviour

Edited by P. F. SECORD (Department of Urban Studies, Queens College, New York) and  
T. MISCHEL (Department of Philosophy, SUNY at Binghamton, New York) assisted by  
R. HARRE (Oxford University)

### CONTENTS

JOSEPH MARGOLIS  
Reductionism and Ontological Aspects of  
Consciousness

HUGH M. LACEY  
The Scientific Study of Linguistic Behaviour

GUY E. SWANSON  
The Primary Process of Groups. Its Systematics  
and Representation

BENG-HUAT CHUA  
On Roles, Rules and Interpretative Understanding

NICHOLAS ABERCROMBIE  
Sociological Indexicality

MARGARET MASTERMAN  
Falling through the Grid or What has Happened  
to the Scarce Women Academics?

Annual Subscription (two issues) £3.50 (\$10.50) Single Issues £1.80 (\$5.40)

Published by BASIL BLACKWELL, Orders to 108 Cowley Road, Oxford OX4 1JF, England

# JOURNAL OF PHILOSOPHICAL LOGIC

*Editor-in-Chief:* Richmond H. Thomason, University of Pittsburgh, Pittsburgh, Pa., U.S.A.

*Editorial Board:* A. R. Anderson, Pittsburgh, Pa.; L. Aqvist, Uppsala; N. D. Belnap, Jr., Pittsburgh, Pa.; M. J. Cresswell, Wellington; J. E. Fenstad, Oslo; F. Fitch, New Haven, Conn.; D. Føllesdal, Oslo; P. T. Geach, Leeds; C. Hamblin, Kensington, N.S.W.; H. G. Herzberger, Toronto; J. Hintikka, Helsinki; W. Hodges, London; S. Kanger, Uppsala; D. Kaplan, Los Angeles, Calif.; S. Kripke, New York, N.Y.; K. Lambert, Irvine, Calif.; R. B. Marcus, Evanston, Ill.; G. J. Massey, Pittsburgh, Pa.; D. Prawitz, Stockholm; N. Rescher, Pittsburgh, Pa.; A. Robinson, New Haven, Conn.; W. C. Salmon, Tucson, Ari.; D. Scott, Oxford; K. Segerberg, Åbo; P. F. Strawson, Oxford; P. Suppes, Stanford, Calif.; R. H. Thomason, Pittsburgh, Pa.; G. H. von Wright, Helsinki.

*Contents of Volume 3, No. 1/2—January/April 1974*

Editorial Announcement.—Ernest W. Adams / The Logic of 'Almost All'.—R. I. Goldblatt / Semantic Analysis of Orthologic.—Ruth Manor / A Semantic Analysis of Conditional Assertion.—Robert K. Meyer / New Axiomatics for Relevant Logics, I.—Francis Jeffry Pelletier / On Some Proposals for the Semantics of Mass Nouns.—James W. Garson / The Substitution Interpretation in Topological Logic.—Michael D. Resnik / On the Philosophical Significance of Consistency Proofs.

Discussions: Relative Truth and Semantic Categories (A Reply to J. F. Post) (Robert L. Martin).—Shades of Possibility (Rejoinder to R. L. Martin) (John F. Post).—Sortal Ranges for Complex Predicates (Robert L. Martin).

Subscription price per volume of four issues (approx. 500 pages), Dfl. 90,—, plus Dfl. 12,50 for postage and handling. One volume will be published yearly. Personal subscription price on request.

**D. REIDEL PUBLISHING COMPANY**

Dordrecht—Holland / Boston—U.S.A.

## The Philosophical Quarterly

CONTENTS OF VOLUME 24      NO. 95      APRIL 1974

*Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant and Stewart McNeilly.  
Published for the Scots Philosophical Club by the University of St Andrews*

### ARTICLES

- I 'GREATNESS' IN ANSELM'S  
ONTOLOGICAL ARGUMENT      R. Brecher
- II EXPLICIT PERFORMATIVE  
UTTERANCES AND STATEMENTS  
   Jane Heal
- III LOGIC AND GRAMMAR      B. H. Slater
- IV MALCOLM ON AFTER-IMAGES  
   P. L. McKee
- V WOLFF'S DEFENCE OF  
PHILOSOPHICAL ANARCHISM Rex Martin

### DISCUSSIONS

- IRVING THALBERG'S COMPONENT  
ANALYSIS OF EMOTION AND  
ACTION      Mitchell Staude
- ON A THEOLOGICAL ARGUMENT  
FOR FATALISM      Susan Haack

### CRITICAL STUDY

- TOULMIN'S EVOLUTIONARY  
EPISTEMOLOGY      Larry Briskman

### BOOK REVIEWS

### LIST OF BOOKS RECEIVED

Annual Subscriptions: Individuals £3.00 (\$8.50) Institutions £5.00 (\$14.00)  
Single Issues £1.00 or \$3.00 (Institutions £1.50 or \$4.00)

for orders placed directly with the Journals Manager, Scottish Academic Press  
Ltd., 25 Perth St., Edinburgh EH3 5DW, Scotland

# Studies in History and Philosophy of Science

**Studies in History and Philosophy of Science** is a quarterly journal which analyses the genesis, evolution and logic of scientific thought. It attempts to unite historical and philosophical approaches to science.

## Recent papers include:

**Maxwell's Methodology and his application of it to Electromagnetism**, A. P. Chalmers. **The Astronomy of Eudoxus: Geometry or Physics?** Larry Wright. **Conceptual Structures and Scientific Change**, Garry Gutting. **Whewell's Theory of Scientific Language**, Morton L. Schagrin. **Speculation in Physics: The theory and practice of Naturphilosophie**, Barry Gower. **Feyerabend and Galileo: the interaction of theories, and the reinterpretation of experience**, Peter K. Machamer. Edited by Gerd Buchdahl, Reader in History and Philosophy of Science, University of Cambridge and Professor L. L. Laudan, Chairman, History and Philosophy of Science, University of Pittsburgh. Published by Macmillan Journals Limited.

**Annual Subscription £8.50 (£9.50 U.S.A. and Canada).** Prices applicable only to orders starting before 31st December, 1974. Payment may be made in any currency at the current exchange rate. Orders must be accompanied by remittance. Cheques should be made payable to: Macmillan Journals Limited, Subscription Department, Brunel Road, Basingstoke, Hampshire, RG21 2XS, England.

## INQUIRY An Interdisciplinary Journal of Philosophy and the Social Sciences

Edited by ARNE NÆSS and ALASTAIR HANNAY

Number 2

Summer 1974

Volume 17

Do We Really Want a *Moral* Justification of Our Basic Ideals?

James R. Flynn

On Being Human

Robin Attfield

Can Philosophy be Original?

T. E. Burke

Objectivism and the Study of Man (Part I)

Hans Skjervheim

### Discussions:

Capitalism and the Desire for Private Gain

Thomas S. Torrance

Reply to Torrance

Richard Schmitt

### Review Discussions:

The Ideology of Philosophy (R. J. Bernstein, *Praxis and Action: Contemporary Philosophies of Human Activity*)

R. A. Sharpe

Humanistic Social Science (M. Roche, *Phenomenology, Language and the Social Sciences*)

H. P. Rickman

Published quarterly by Universitetsforlaget, P.O. Box 307, Blindern, Oslo 3, Norway

Annual Subscription (Postage included): N. kr. 75.00, £4.90, \$12.50 (Institutions)

N. kr. 60.00, £3.90, \$10.90 (Individuals)

# **PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH**

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

in cooperation with a distinguished group of American and Foreign Scholars

Descriptive, analytic, critical, and historical articles representing the major contemporary trends in philosophy. In addition to phenomenological research, PPR publishes studies in a wide range of areas including ethics and value theory, metaphysics, aesthetics, logic, language, political, social and religious philosophy, theory of knowledge, and the sciences.

Address all communications to  
*Philosophy and Phenomenological Research*  
State University of New York at Buffalo  
Buffalo, New York 14226

Annual subscription rate, \$10.00 for libraries and institutions, \$8.00 for individuals; single copies \$2.50 and \$2.00 respectively, plus postage and handling on back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

Vol. 49 No.188

## **Philosophy**

April 1974

The Journal of the Royal Institute of Philosophy

The editorial policy of the journal pursues the aims of the institute: to promote the study of philosophy in all its branches: logic, metaphysics, epistemology, ethics, aesthetics, social and political philosophy and the philosophies of religion, science, history, language, mind and education.

### **Contents**

Editorial: **Uri Geller**

The Conscience of Huckleberry Finn **Jonathan Bennett**

Law, Liberty and Indecency **David A. Conway**

The Concepts of the Sceptic: Transcendental Arguments and Other Minds **G. W. Smith**

The Positive McTaggart on Time **John King-Farlow**

Saints, Heroes and Utilitarians **Christopher New**

### **Discussion**

Wittgenstein and Heraclitus: Two River-Images **Roger A. Shiner**

Explanatory Completeness **Donald Scherer**

A Lack of Discipline **R. A. Becher**

Philosophy is published quarterly by Macmillan Journals Ltd., 4 Little Essex Street, London WC2R 3LF. Annual subscription £4.00 (£5.00 U.S.A. & Canada).

Edited by  
D. W. HAMLYN

*Articles*

Expressions of Intentions:

D. GUSTAFSON

Describing, Evaluating, and Moral

Conclusions: C. M. CHERRY

Retained Knowledge: A. HOLLAND

Backwards Causation and Continuing:

S. WATERLOW

On Our Knowledge of Matters of Fact:

E. SOSA

Criteria for Plausible Arguments: J. AGASSI

*Discussions*

Grice and MacKay on Meaning:

R. T. GARNER

On 'Levels of Rules and Hart's Concept of

Law': T. M. BENDITT

Negative Utilitarianism: A. D. M. WALKER

The Explanatory Power of Chomsky's Trans-

formational Generative Grammar:

J. SINGLETON

Book reviews

Freedom as the Non-Restriction of Options:

W. A. PARENT

Freedom as the Non-Restriction of Options: A

Rejoinder: S. I. BENN and W. L.

WEINSTEIN

Nullity and Sanction: P. MULLOCK

Tractatus 5.542: M. COHEN

Rights Without Justice: T. D. CAMPBELL

Books Received

Published for the MIND ASSOCIATION by BASIL BLACKWELL

Single Issues: 80p

Annual Subscription: £2.50

## Theoria A Swedish Journal of Philosophy

Edited by GÖRAN HERMERÉN and KRISTER SEGERBERG

---

### Special Issue on Aesthetics

Volume XXXIX

CONTENTS

1973, 1-3, 192 pp.

- |                     |  |
|---------------------|--|
| ARTHUR C. DANTO     | Artworks and real things   |
| RICHARD J. SCLAFANI | Artworks, art theory, and the artworld   |
| ALAN TORMEY         | Critical judgments   |
| MONROE C. BEARDSLEY | What is an aesthetic quality?  |
| GÖRAN HERMERÉN      | Aesthetic qualities, value and emotive meaning                                       |
| MORRIS WEITZ        | Interpretation and the visual arts   |
| TED COHEN           | Aesthetic/Non-aesthetic and the concept of taste:<br>a critique of Sibley's position |
| GEORGE DICKI        | Taste and attitude: the origin of the aesthetic                                      |
- 

THEORIA, which invites contributions in all fields of analytic philosophy, is published three times a year. Subscriptions and requests for back numbers should be sent to the publishers, Messrs. G. W. K. Gleerup, Box 1205, 221 05 Lund 1, Sweden.

Annual subscription Skr. 35: (approx. \$7 US).

ISSN 0040-5825



# AMERICAN PHILOSOPHICAL QUARTERLY MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

---

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

---

- No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$7.00 (£2.90).
- No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$7.00 (£2.90).
- No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$7.00 (£2.90).
- No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$7.00 (£2.90).
- No. 5. DESCARTES' PHILOSOPHY OF NATURE, by James Collins. 1971, \$7.00 (£2.90).
- No. 6. STUDIES IN THE PHILOSOPHY OF MIND. *Contents:* Harold Brown, "Perception and Meaning"; Haskell Fain and A. P. Griffiths, "On Falsely Believing That One Doesn't Know"; O. H. Green, "Emotions and Belief"; Donald F. Henze, "Descartes on Other Minds"; Moreland Perkins, "Meaning and Feeling"; and Alan R. White, "What We Believe." 1972, \$7.00 (£2.90).
- No. 7. STUDIES IN ETHICS. *Contents:* Charles V. Blatz, "Mitigating and Meliorating Defenses"; Claudia Card, "Retributive Penal Liability"; Norman O. Dahl, "Is Mill's Hedonism Inconsistent?"; R. L. Franklin, "Recent Work on Ethical Naturalism"; James D. Wallace, "Cowardice and Courage"; A. D. Woozley, "Injustice." 1973, \$7.00 (£2.90).
- No. 8. STUDIES IN MODALITY, by Nicholas Rescher, with the collaboration of Ruth Manor, Arnold vander Nat, and Zane Parks. 1974, \$7.00 (£2.90).

# AMERICAN PHILOSOPHICAL QUARTERLY

Edited by  
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

|                      |                    |                    |
|----------------------|--------------------|--------------------|
| • Virgil C. Aldrich  | James Collins      | Hughes Leblanc     |
| Annette Baier        | Antonio S. Cua     | Keith Lehrer       |
| Kurt Baier           | Arthur C. Danto    | Gerald J. Massey   |
| Stephen F. Barker    | Richard M. Gale    | George Mavrodes    |
| Monroe Beardsley     | Peter Thomas Geach | Richard Rorty      |
| Nuel D. Belnap, Jr.  | Marjorie Grene     | George A. Schrader |
| Myles Brand          | Adolf Grünbaum     | Wilfrid Sellars    |
| Robert Butts         | Rom Harré          | Ernest Sosa        |
| Roderick M. Chisholm | Carl G. Hempel     | Richard Taylor     |
| L. Jonathan Cohen    |                    | John W. Yolton     |

VOLUME 11/NUMBER 4

OCTOBER, 1974

## CONTENTS

|   |     |   |     |
|---|-----|---|-----|
| I. EDWIN CURLEY: <i>Recent Work on 17th Century Continental Philosophy</i> . . . . .                          | 235 | V. SOUREN TEGHRARIAN: <i>Linguistic Rules and Semantic Interpretation</i> . . . . . | 307 |
| II. PETER ACHINSTEIN: <i>The Identity of Properties</i> . . . . .   | 257 | VI. VIRGIL C. ALDRICH: <i>Sight and Light</i> . . . . .                             | 317 |
| III. ROBERT C. SOLOMON: <i>Hegel's Epistemology</i> . . . . .   | 277 | VII. ROBERT HOFFMAN: <i>Vulgarity</i> . . . . .                                     | 323 |
| IV. RICHARD SCHACHT: <i>On "Existentialism," Existenz-Philosophy and Philosophical Anthropology</i> . . . . . | 291 | VIII. JOHN BEVERSLUIS: <i>Socratic Definition</i> . . . . .                         | 331 |
|   |     | IX. DALE GOTTLIEB: <i>Foundations of Logical Theory</i> . . . . .                   | 337 |
|   |     | <i>Books received</i> . . . . .   | 344 |
|   |     | <i>Corrigenda</i> . . . . .   | 345 |

# AMERICAN PHILOSOPHICAL QUARTERLY

## POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

## MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

## COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

## OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$14.00, increasing by \$1.40 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$11.20, increasing by \$1.40 for each additional 4 pages. Covers will be provided for offprints at a cost of \$6 per group of 50.

## SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers (including the annual Monograph) and eighteen dollars for institutions (excluding the annual Monograph which is obtainable on payment of another seven dollars). Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and five dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

## MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



# I. RECENT WORK ON 17TH CENTURY CONTINENTAL PHILOSOPHY

EDWIN CURLEY

THE decade just past has been a fruitful and exciting one for those interested in the history of continental philosophy in the 17th century. Not only has there been a great deal of work published on that period, but an encouragingly high proportion of it has proven to be well worth publishing. We have good, new translations and critical editions of many works of both major and minor figures, new bibliographical guides to the literature on them, and a growing body of instructive secondary works. The history of philosophy seems now to be attracting, with much greater frequency than was the case, say, 20 years ago, writers who combine a sympathetic and scholarly approach to the philosophers they study with genuine philosophical competence of their own in sorting out the issues raised and saying something about those issues. While there is still much that needs to be done to improve our understanding of the period, the record of the past decade has been a hopeful one.

Under these circumstances, one can hardly expect to do more than give a brief indication of some of the highlights of recent work. I shall concentrate on studies of the philosophers whom I regard as major and central—Descartes, Spinoza, Malebranche, and Leibniz—leaving it primarily to the bibliography to provide some guide to the often very interesting work on philosophers who must be regarded as comparatively minor or peripheral (e.g. Bayle, Gassendi, Galileo, and Pascal), and to such aids to scholarship as translations, editions, and bibliographies. Aids to scholarship which are of special interest will be starred items in the bibliography.

## I. DESCARTES

Recent English work on Descartes has concentrated on certain classical cruxes of interpretation—the *cogito*, the circle, and the ontological argument—and has done a great deal to clarify the problems these topics involve. A good example is provided by Hintikka's article on the *cogito* (6).

It is doubtful whether many of the positive suggestions Hintikka makes will survive critical scrutiny, but there is no doubt that the discussion he provoked has been extremely helpful.

The puzzle with which Hintikka is primarily concerned is this: Does Descartes regard his knowledge of his own existence as intuitive or as based on an inference? Both alternatives can be supported textually and both alternatives present philosophical difficulties for the Cartesian system. Descartes does say in the *Regulae* (Rule 3) that everyone sees by an intuition that he exists, that he thinks, and many other things. And in the *Second Replies* he seems to say again that his existence is known intuitively. But this approach invites embarrassing questions about the reliability of our intuitions and about Descartes' right to assume their reliability at the stage in the *Meditations* when he puts the *cogito* forward. Again, in most of Descartes' presentations of the *cogito* he certainly makes it look as though he is inferring existence from thinking. Except for the *Regulae*, which is an early work, and the *Second Meditation*, the standard formula tends to be "I think, *therefore*, I exist." But how is this compatible with the apparent status of "I exist" as a first principle? And how would Descartes justify his reliance on the premisses which, *prima facie*, are involved?

Hintikka offers us the notion of "I exist" as something like a performative utterance, rather than an inference. Contrary to the suggestion of his title, he does not regard the alternative he sets before us as an exclusive one. We need not choose between inference and performance. The *cogito* is both. But Hintikka prefers to emphasize the performative aspect, partly because he thinks that, viewed as an inference, it is not a very good argument.

As an inference, Hintikka contends, the *cogito* is an instance of *modus ponens*, with a suppressed conditional premiss of the form:

$$(i) \quad Ba \supset (\exists x)(x=a)$$

While Hintikka concedes that this conditional is

provable in standard systems of predicate logic, he thinks that in the sense in which it is true it is of no use to Descartes in proving his conclusion. The systems of logic in which (i) is provable all tacitly assume that the singular terms employed really refer to some actually existing individual. If we make this assumption, then before we can assert the categorical premiss of our argument,

(ii)  $Ba$

the "I think," we must be in a position to assert the conclusion,

(iii)  $(\exists x)(x=a)$

the "I exist." The argument will be formally valid and its premisses will be true, but it will be question-begging, since establishing the truth of the conclusion will be a necessary condition of establishing the truth of the categorical premiss.

But if we drop the assumption that all singular terms designate actually existing individuals, then, though we shall be able to determine the truth of our categorical premiss without first determining the truth of our conclusion, the conditional premiss will be false.  $Ba$  will be true where  $Bx$  is read " $x$  thinks" and  $a$  designates Hamlet, but  $(\exists x)(x=a)$  will be false for that interpretation of  $a$  and so  $Ba \supset (\exists x)(x=a)$  will not be a logically true formula. Hintikka has other, textual objections to reading the *cogito* as an inference, but he is well aware that the textual evidence points both ways. This is the difficulty he treats as most serious.

So he proposes to regard the *cogito* as a performance, though this is held to be only the better half of Descartes' insight. Descartes should be viewed as recognizing the absurdity of performing a certain kind of action, denying his own existence. The sentence "I don't exist" though formally consistent, and therefore not logically false, is nevertheless *existentially* inconsistent, and hence, impossible to defend or believe. Its contradictory, "I exist," is therefore self-verifying. This is a sophisticated modern version of the view that "I exist" is certain on intuitive grounds.

A sentence,  $p$ , is existentially inconsistent for  $a$  to utter (assertively) if and only " $p$  and  $a$  exists" is inconsistent (in the ordinary sense). E.g., "Nixon does not exist" is existentially inconsistent for Nixon to utter. Hintikka speaks of the utterance of a sentence by a person for whom the utterance of that sentence is existentially inconsistent as an *existentially inconsistent statement*.

Normally an existentially inconsistent statement will be pointless, since it will defeat one of the

standard purposes of uttering a declarative sentence, that of making the hearer believe what one says. If the hearer identifies the man speaking as the man the sentence is about, he will recognize that the statement is false. But the hearer might understand the sentence without identifying the speaker as the man the sentence is about. Hence, existentially inconsistent statements are not inevitably self-defeating.

"I do not exist" is a special case in two respects. First, since the function of 'I' is simply to refer to the speaker, the sentence is existentially inconsistent for anyone to utter. Secondly, here it is a necessary condition of the hearer's understanding the sentence that he identify the speaker as the man spoken about. Unless he does make the identification, he does not understand the use of 'I'. So "I don't exist" is inevitably self-defeating for anyone to utter assertively, and therefore, absurd in a very special way. Conversely, "I exist" is self-verifying.

One problem for any interpretation along these lines is to explain why Descartes so often gives the illusion that he is inferring his existence from his thought, when really he is recognizing the self-evidence of his existence. What are the words *cogito* and *ergo* doing in the *cogito ergo sum*?

Hintikka's claim is that the "I think" expresses the performatory character of Descartes' insight. It is from an act of thought that Descartes comes to recognize the indubitability of his own existence, viz., the act of trying to think the contrary. Descartes cannot *think* that he does not exist, in the sense of making himself believe it. The attempt to make himself believe that he does not exist is necessarily self-defeating. So the relation of *cogito* to *sum* is more a causal one than a relation of inference. A certain (attempted) thought act necessarily produces a conviction of existence. It would be more accurate for Descartes to say "By thinking I perceive my existence." But the thinking part is quite essential. Not just any action will do. It must be an act of thought, specifically, an act of trying to persuade myself that I do not exist.

Hintikka's interpretation has a certain philosophic interest in its own right, whatever its value may be for the understanding of Descartes. His remarks about the logical peculiarities of the sentence "I do not exist" strike me as correct and important. But his interpretation encounters some difficulties as an account of Descartes' thought, difficulties which I think have been brought out best in an article by Harry Frankfurt (5).

His criticism centers on Hintikka's explanation of the role of thinking in the *cogito*. There are two main points. First, Hintikka's interpretation requires that the thought act which results in my perception of my own existence have a very specific content: what I am trying to think is that I do not exist. In failing to believe this, I realize that I do exist. But the thought acts with which Descartes is willing to connect his own existence are quite varied in their content—my existence follows from my being persuaded that there are no bodies, no heaven, no earth, etc.; it follows from my judging that the wax exists or from my judging that I do exist. Secondly, Hintikka's interpretation requires that the thought act be of a very special kind. It will not do for the performative interpretation if I simply contemplate the possibility of my not existing, for that is something I certainly can do. The thought act must be an attempt to think my present non-existence with assent. But for Descartes it is clear that a wide variety of thought acts will do. I can infer my existence from my doubting or even from the fact that I entertain the proposition "I exist."

Hintikka, of course, can not be decisively refuted by such evidence. He was familiar with the relevant passages when he first wrote, and his insistence that Descartes did not keep consistently to any one interpretation renders him virtually immune to textual rebuttal. What is needed is a satisfactory answer to his philosophic objections to treating the *cogito* as an inference.<sup>1</sup> And I think the work of Popkin (34) and Frankfurt (4) does point the way to such an answer. What follows is an attempt to outline a possible response. It is not so much an account of what has been said, as a variation on certain prominent themes of recent studies.

Hintikka's central objection seems to be this: If we make certain assumptions about the singular terms we use, then "I think" is logically sufficient for "I exist." But then knowing that "I exist" will be a necessary condition of knowing that I think. And so the inference will not be probative.

One might be tempted to reply that if so, then any valid deductive argument would have to fail as a proof. For if it is valid, the truth of the premisses will be sufficient for the truth of the conclusion and the way will be open for someone to argue that knowing the truth of the conclusion is a necessary condition of knowing the truth of the premiss.

Hintikka's objection to the *cogito* begins now to resemble Mill's objection to the syllogism, and to be applicable to deductive reasoning generally. Many might think this a sufficient response. It is patently false that all deductive reasoning involves a *petitio*.

But would Descartes feel that he could take it for granted that some deductive reasoning is probative? Criticisms of the syllogism like Mill's had been made by the sceptics as early as Sextus Empiricus, repeated by Renaissance writers like Sanchez (see Popkin (34)) and apparently accepted by Descartes himself in the *Regulae*. Would he not have thought that the accusation of question-begging raised a serious problem?

The correct reply to Hintikka's objection, I suggest, is that while the *cogito* is an inference, it is not a demonstration or proof in any classical sense. It is not essential to an inference that the premisses of the inference be known to be true, though this is essential to a proof. It is because we think the *cogito* must be a proof, if it is an inference at all, that we think its premiss must be known (cf. Kenny (8)). But this only obscures the point of Descartes' enterprise.

As Popkin and Frankfurt have stressed, Descartes is fundamentally concerned with overcoming the difficulties raised against "dogmatic" philosophers by the sceptics. I take it that he is particularly anxious about the objections they made against the notion of proof. Proof cannot go on to infinity. The process of establishing one proposition by citing others which entail it and are known to be true must come to an end somewhere in propositions which are not themselves proven. The problem is to find a premiss or set of premisses which can serve as first principles without reasonably incurring the accusation that, since they are not proven, they are merely arbitrary assumptions.

Descartes does not want to rely on the contention that his first principles are self-evident. He knows this would be dismissed by the sceptic as a bare assertion, a matter of subjective conviction, liable to great variation from one person to another. So he adopts a negative procedure for getting his first principles. As Frankfurt puts it, "he establishes truths by removing the grounds for doubting them rather than by proving their truth in a direct way" ((4), p. 174).

Cartesian doubt must always be *reasonable* doubt,

<sup>1</sup> The apparent textual counter-evidence is very well treated in Bernard Williams' article in the Doney collection (3). See also Weinberg (11), Kenny (8) and Röd (10).

but the requirements for reasonable doubt are deliberately very weak. The ground offered for doubting need not have anything positive to be said for it, though in general there should be no intellectually compelling argument against it. Descartes' procedure for "establishing" first principles is to accept all and only those propositions which he can conceive no reasonable ground for doubting. He will systematically apply to his former beliefs the most powerful grounds of doubt he can find. What survives the procedure cannot be said to be accepted arbitrarily. To believe only what you have no reasonable ground for doubting—in this weak sense of "reasonable"—is to have very high standards of rationality. But to accept what survives systematic doubt is not to put something forward as the conclusion of a proof.

That I exist is, as Descartes insists, a first principle. But it is an acceptable principle because it survives systematic doubt. Any ground for doubting my existence which purports to be reasonable will have to explain my erroneous belief in my existence. So it will have to involve the supposition that I think. And thought entails a thinker. So "I exist" is inferred from "I think," but "I think" is not a premiss of a proof, not a proposition which Descartes must take any responsibility for defending as known to be true. It is an essential element in any hypothesis which I might entertain, or an opponent might suggest, to throw reasonable doubt on my existence.

The emphasis placed here on the requirement that doubt be reasonable is also important in Frankfurt's proposed solution to the problem of the circle (4). Suppose we put this difficulty as Arnauld put it—"we are certain that the things we conceive clearly and distinctly are true because God exists . . . we can only be certain that God exists because we conceive that very clearly and distinctly, so before we can be certain that God exists, we must be certain that whatever we conceive clearly and distinctly is true."

Frankfurt's reply is roughly as follows. If being certain is equivalent to being unable to doubt, then it is just a mistake to suppose that we can be certain of God's existence only if we already are certain that whatever we conceive clearly and distinctly is true. So long as we are attending to the proof of God's existence and conceiving it clearly and distinctly, we shall be in a state where *non possumus non credere*. All present clear and distinct ideas compel assent and are, in that sense, indubitable.

But indubitability, in that sense, is not equivalent

to knowing. It is a psychological state of complete conviction, from which we may fall when we are no longer attending to the arguments which induced it. "Indubitable" can, however, be given a normative sense. Though it may be possible to doubt propositions which were once, but are not now, conceived clearly and distinctly, it may not be possible to doubt them reasonably. In the *First Meditation* the supposition that there may be a deceiving God constitutes a reasonable ground for doubting the truth of the things that seem most evident to us. But by the end of the Fourth Meditation it does not since there is then available a compelling argument against this supposition and none in its favor. Frankfurt construes Descartes as concerned to counter the sceptic's claim that the use of reason leads to the conclusion that reason is unreliable, that there are good reasons for being suspicious of the value of good reasons. The Cartesian reply is that a scrupulous use of reason leads rather to the conclusion that reason is reliable.

Such, in its broad outlines, is Frankfurt's reconstruction of the Cartesian position. So far, the reconstruction strikes me as both correct and very illuminating. Indeed, I would be more sanguine than Frankfurt is about the ability of this approach to deal with the philosophical difficulty. But there is a good deal in the detailed development of Frankfurt's position, particularly in (4), which is open to objection.

Frankfurt allows that Descartes' conclusion—that it would never be reasonable to doubt what we perceive clearly and distinctly—may be true without its being the case that all our clear and distinct ideas are true. And he cites a passage from the *Second Replies* in which Descartes appears to admit the possibility that what we intuit is false even though we have no reasonable grounds for suspecting error. He interprets Descartes as being more concerned with the solidity and permanence of his beliefs than he is with their "truth" in any absolute sense. The logic of Descartes' position leads to a coherence, rather than a correspondence, theory of truth, and it is the coherence of our carefully adopted beliefs which God's veracity guarantees. Applying this to the question of the status of scientific theories, a vital problem for Descartes and his contemporaries because of the controversy between the "two great systems" of astronomy, Frankfurt suggests that the *Meditations* imply a Kantian solution to the problem: "men may content themselves with certainty about the phenomena and leave the noumena to God."

This is a strange conclusion for an author who has emphasized Descartes' role as the opponent of scepticism. Popkin (34) cites the same passage from the *Second Replies* and reads it in the same manner. But he treats it as a critical and damaging concession to scepticism. And so it seems to be as Popkin and Frankfurt read it. This gives us a good reason, I think, for seeking an alternative interpretation of the *Second Replies*. Further grounds are provided by the *Principles*, where Descartes takes up an explicit position on the Copernican controversy which cannot, I think, be reconciled with the position Frankfurt finds implicit in the *Meditations*. But this issue is too complex to be argued here.

If Hintikka, Popkin, and Frankfurt have done much for our understanding of the *cogito* and the circle, perhaps the most important contribution of Kenny's excellent book on Descartes (8) lies in his discussion of the ontological argument. The classic objection to the argument, raised, in effect, by Cassendi, well before Kant—is that existence is not a property. What seems to lie behind this objection (cf. Alston (1)) is the view that the existence of a subject of predication is a necessary condition for predicating any property, either truly or falsely, of the subject. If that view is correct, and if existence is a property, then we can never make false positive existential statements or true negative ones. But this is patently false. Thus, existence is not a property.

Kenny argues that Descartes' reply would be to reject the view that true or false predications of properties require an existing subject. We have *a priori* knowledge of, and therefore can make true predications about, the objects studied in geometry. It does not matter that there are not now, never have been, and never will be any such objects. All that is required is that the objects be possible objects. So Kenny interprets Descartes as being committed to something like Meinongian pure objects as subjects of predication. These objects are not to be regarded in a conceptualistic fashion. Their *esse* is not *cogitari*. Rather they impose a necessity on thought.

Kenny does not claim that this is, in the end, a satisfactory way of rebutting the classical objection. Though he holds that Descartes' version of the theory of pure objects avoids some of the difficulties Meinong's encounters, he thinks there are insurmountable difficulties in the notion of non-existent possible individuals, and cites Quine's contention that we lack satisfactory identity conditions for pure objects.

Moreover, Kenny maintains Descartes' answer gives rise to serious internal difficulties in the Cartesian system. If the ontological argument is not to be circular, it must be possible to prove properties of the problematically existent. But if so, then what happens to the *cogito*? If "This triangle has its three angles equal to two right angles, so this triangle exists" is not valid, then the inference, "I think, so I exist" cannot be valid. Conversely, if the inference from thought to existence is valid, then Descartes cannot prove that the triangle (or God) has certain properties without first knowing whether the triangle exists. Kenny concludes that the *cogito* and the ontological argument cannot both be valid, which he regards as a major problem for the Cartesian system.

As far as the interpretation of Descartes goes, I think Kenny is quite right to attribute a theory of possible objects to Descartes. This is argued in more detail in (9) than in (8) and it seems to me to be supported even more strongly by a passage Kenny does not cite (AT V, 160). But it is not clear that the theory cannot be defended for some distance against Quinean objections, or that it is inconsistent with the *cogito*. Presumably one wants to make the subject predicate distinction both with actual objects and with possible objects. As Kenny himself remarks, though "neither Pegasus nor Pegasus' shape exists . . . Pegasus is a non-existent substance, whereas Pegasus' shape is a non-existent mode" (7). The crucial question is whether the thinking which is posited in the premiss of the *cogito* is posited as an existent mode. If it is, then it will require an existent substance to inhere in.

As for the queries about whether the possible fat man in the doorway is identical with the possible bald man in the doorway, it might be argued that these descriptions are not sufficiently definite for us to be able to answer identity-questions about their referents (Cf. Rescher, (35)). Is the Bronte sister who wrote novels identical with the British authoress who was a younger sister? How could we possibly decide? Neither description picks out just one individual, even if we restrict ourselves to the actual. If we allow our descriptions to range over non-existent possible individuals, then no putatively definite description will pick out just one individual unless it is a complete description in the Leibnizian sense. So we should not expect to be able to answer, or even satisfactorily frame, specific identity questions about possible individuals, though we might be able to say that in general two possible



individuals are identical if their individual concepts are identical. In the end the difficulties the ontological argument involves concern the principle of the identity of indiscernibles and the notion of a complete individual concept. So the confrontation of Descartes with 20th century objections leads very quickly to reflection on the problems of the Leibnizian system.

## II. SPINOZA

In the recent literature on Descartes, there is a considerable body of works whose virtues are as much philosophical as they are scholarly in the sense that these works offer as much to interest someone for whom philosophy effectively begins in this century as they do to interest one whose primary concern is with what Descartes' views really were. There is not yet a comparable body of philosophical scholarship on Spinoza, whose study remains more the preserve of historians who are not at the same time deeply involved in contemporary debates. There is, nonetheless, a great deal of interesting recent work, the bulk of which is French, and the most significant of which, clearly, is Martial Gueroult's massive three volume study of the *Ethics* (33). So far only the first volume, dealing with Part I of the *Ethics*, has appeared. But it is already plain that the work will be of immense value to students of Spinoza. In his nearly 600 pages of commentary on the 39 pages of Spinoza's text, there is hardly any important problem of interpretation which he does not discuss, and in some measure at least, illuminate.

The central theme of Gueroult's work, which guides both his analysis of the structure of *De Deo* and his interpretation of its doctrines, is Spinoza's rationalism. Rationalism is understood here, not as a doctrine about the possibility of synthetic a priori knowledge, but as an affirmation of the complete intelligibility for man of the essences of all things, both finite and infinite. Gueroult rightly makes this the primary article of faith for Spinozism, and he draws from it a number of important conclusions, of which I shall comment on two which I find particularly interesting.

The first concerns the relation of the attributes to substance. Gueroult's stress on the absolute intelligibility of God in Spinoza's system leads him to reject strongly the "subjectivist" interpretation of the attributes according to which the attributes have no real existence in the essence of God, but are only "subjective modes of thinking," ways in

which the mind conceives a simple reality which lies forever beyond its grasp, like a Kantian noumenon. Gueroult devotes a long and (to my mind) entirely successful appendix to a detailed refutation of the case for this widespread misinterpretation.

Gueroult's own interpretation conceives Spinoza's substance as a complex whole whose essence is constituted by an infinity of really distinct elements, the attributes, each of which satisfies in its own right the definition of substance. So substance is identical with the totality of its attributes, and each attribute both exists in itself and is conceived through itself. This is very similar to a position I have argued for myself (31), but Gueroult is able to support it with textual evidence I missed (e.g., E IP29S which says quite explicitly that the attributes exist in themselves) and takes up problems inherent in it which I neglected. As Gueroult points out, a good many objections to Spinozism can be shown to rest on a misunderstanding if this interpretation is accepted (p. 120). But it is not without difficulties of its own.

If each of the attributes is really distinct from every other attribute and if each satisfies the definition of substance, then it is not easy to see why we should not say that the attributes constitute an infinity of distinct substances rather than one substance. This is a central objection to any interpretation like Gueroult's and providing an answer to it is one of his main preoccupations in the first half of his commentary. If I understand him, his reply is this. If we considered the attributes separately, in abstraction from the concept of God as an absolutely infinite being, then it would be perfectly correct to think of each of the attributes as a distinct substance, existing in itself and conceived through itself. But once we conceive the idea of an *ens realissimum*, of a being whose nature it is to consist of infinitely many attributes and whose existence follows from that nature, then we are obliged to conceive of the attributes as constituents of *one* being and not as a mere collection of substances. God's existence does not result from the necessary existence of each of the elements which constitute his nature. It is rather their existence *together* which results from his necessary existence (p. 184)—or, as I think Gueroult would add, from his necessary existence as an indivisible and unique being, the elements of whose nature cannot be conceived as separable from it (p. 220), a being which excludes the existence of any other substance (p. 226).

Whether or not this will be regarded as an

ultimately satisfactory reply to the objection, Gueroult offers good reason for supposing that it was essentially Spinoza's reply. I might add that, given the necessary existence of each of the attributes considered in itself, the question of their existing as a collection of really distinct substances in the Cartesian sense of "real distinction" cannot arise. For though the attributes are not dependent on one another for their existence, nevertheless, since each attribute exists necessarily, it is simply not possible that one should exist without the others. Gueroult would not, I suspect, regard this as being an adequate explanation of their unity in one substance, but it does point up the fact that, even when considered as a collection the attributes are a very special sort of collection.

The second conclusion of Gueroult's which I want to comment on concerns his interpretation of the infinite modes, which seems to me to raise very serious difficulties. Spinoza's rationalism, he argues, requires that the human intellect be regarded, not as an effect of God's intellect, but as a part of it (p. 404). In this way any radical discrepancy between God's conception and man's (such as was contemplated by Descartes) is effectively blocked. The divine intellect, of which the human intellect is a part, does not pertain to *natura naturans* (i.e., the attributes) but to *natura naturata* (i.e., the modes) and is identified by Gueroult with (both) the infinite modes of the attribute of thought. On Gueroult's reading the immediate infinite mode of thought is the totality of thought essences (essences pensées, p. 318), or the universe of eternal ideas, whereas the mediate infinite mode of thought is the totality of finite modes of thought existing in duration. Any existing finite intellect is a part of the mediate infinite mode of thought; its eternal idea is a part of the immediate infinite mode of thought; and these two modes constitute the intellect of God. A similar analysis is applied to the infinite modes of extension.

Gueroult presents a case for this interpretation which is as well-documented as we could reasonably expect, considering how little Spinoza is prepared to tell us about the infinite modes, and it may be that he is right on the point of interpretation. But if he is, then Spinoza's doctrine of the nature of these modes looks inconsistent with what he holds concerning their causation. For as I've argued (31) and as Gueroult would seem to agree (p. 313 *et passim*), Spinoza's doctrine is that the attributes are in themselves adequate causes only of their infinite modes, but are not, in themselves, adequate causes

of their finite modes, whose production requires, as a further condition, the existence of other finite modes. But surely if an attribute is an adequate cause of a whole, it cannot at the same time fail to be an adequate cause of the individual parts of that whole. So if Spinoza does conceive of the finite modes as parts of the infinite modes, then his system is inconsistent.

We might note in this connection that Gueroult makes virtually no use of the notion of the laws of nature in his explication of Spinoza. I have been able to locate only one reference to laws (p. 250), and there they are identified with the series of finite things. No grounds for this are offered, and it does appear to be inconsistent with TdIE, 100-101, where the laws are said to be inscribed in the fixed and eternal things, which are sharply distinguished from the series of singular mutable things. This latter passage may offer the key to a different and more satisfactory interpretation of the infinite modes. But this is too complex an issue to be debated here.

### III. MALEBRANCHE

The most striking thing about recent work on Malebranche is that it is no longer a peculiarly French preoccupation. The first 60 years of this century saw only one full-length study of Malebranche in English and very few in any other non-French language. The past decade alone has added three more in English: Rome's (29), Walton's (30) and Connell's (23), and one in Italian, Nicolosi's (25). Besides these there are substantial French works by Rodis-Lewis (28) and Robinet (26, 27), and the papers collected in *Malebranche: L'Homme et l'oeuvre* (22). We can certainly not complain that Malebranche is being neglected by the scholars, though we might wish that their work would reach a wider audience. Malebranche is too interesting and important a figure to be left to the historians.

One of the principal issues in the flood of materials centers on the continuity of Malebranche's thought. There has been a tendency in French scholarship, aided recently by the new critical edition of the complete works (24), to emphasize the discontinuities, to agree with Arnauld in seeing a very different doctrine of the vision in God after the 10th *Eclaircissement* than was offered in *La recherche de la vérité*. E.g., Robinet's work (27) is a detailed elaboration of five distinct stages in Malebranche's development, with a most important break occurring in what Robinet calls "la crise de

1677." Before 1677, Malebranche conceives the ideas we see in God as particular, finite, created beings, a conception inspired in part by the Cartesian doctrine of the creation of the eternal truths. After 1677, in response to the criticisms of Foucher, Malebranche rejects the doctrine of the creation of the eternal truths and conceives what is seen in God as the unique idea of intelligible extension, infinite, general, and uncreated, but capable of representing the essences of all material things.

By contrast, the tendency of English language scholars, like Rome and Connell, has been to emphasize the continuities, treating the later versions of the doctrine as simply logical developments of the earlier versions. In this they may find encouragement in the work of Mme. Rodis-Lewis, who sees Malebranche's thought as evolving so gradually that the change is almost imperceptible. In particular she has been sceptical of Robinet's notion of a crisis of 1677, questioning whether Malebranche ever accepted the creation of eternal truths. See not only her (28), but also her contribution to (22), followed by a discussion with Gouhier, Robinet, and Gueroult, among others.

#### IV. LEIBNIZ

One of the interesting features of recent work on Leibniz is the extent to which it has been concerned with the reassessment of the critical works of Russell and Couturat, which are now some 70 years old. Though there has been a growing disposition to disagree with Russell's and Couturat's conclusions, nevertheless their work has provided a focus for much that has been written in the past decade. So, e.g., Kauppi (18) argues that Couturat's account of Leibniz' logic was distorted by Couturat's own preference for the extensional point of view, a preference characteristic of the period in which Couturat wrote, but which would not be so widely shared by logicians today. Similarly, Parkinson (20) takes up Russell's claim that Leibniz' metaphysics was almost entirely derived from his logic and subjects it to a very patient examination, culminating in the (perhaps not very surprising) conclusion that Russell's list of Leibniz' premisses is quite defective. Criticism of Russell and Couturat is also prominent in Rescher's work on Leibniz (21) and in the recently published monograph by Hide Ishiguro (17). These are perhaps the most important recent monographs on Leibniz, but as befits a philosopher who is not being left to

the historians, there are also a considerable number of interesting articles on Leibniz, many of which are collected in Frankfurt's anthology in the Anchor series (14).

Out of all this discussion there are a number of areas in which the conclusions of Russell and Couturat are being reversed. One concerns the problem of contingency and the principle of sufficient reason. Couturat had argued, and had persuaded Russell, whose earlier views were different, that the principle of sufficient reason was given by the formula: all true propositions are analytic. Couturat would not allow any exceptions to this doctrine for existential propositions or any place in Leibniz' philosophy for final causes. The only difference between necessary and contingent truths was that the analysis of contingent truths involved an infinite process which surpassed the powers of the human mind. So the system had a distinctly Spinozistic tendency.

Recent discussions by Parkinson (20), Rescher (21), and myself (14) have tended in some measure to go back to Russell's original view (before Couturat) that, apart from the proposition that God exists, existential truths are not analytic. Though both Parkinson (p. 106) and Rescher (p. 25) do, in various places, ascribe to Leibniz the view that all truths are analytic, they both also make an exception for existential truths about finite individuals (Parkinson, p. 107; Rescher, p. 18). And to the extent that they identify the principle of sufficient reason with the doctrine that all truths are analytic, they argue that it needs to be supplemented by a further principle—the principle of perfection or the best, which turns out to involve the assertion that God, in creating the world, chose to act in the best possible way. So final causes are, in the end, re-instated.

But very divergent views have been held about the status of this further principle. Parkinson argues (p. 109) that according to Leibniz, it is necessary (absolutely, not just hypothetically) that God should choose the best. Parkinson does not think this excludes the contingency of the world, since other worlds remain possible (i.e. free from contradiction in their own nature) even if they are not possible in relation to God's will. I, on the other hand, have argued (p. 95) that God's choice of the best does not follow from his nature, and Rescher has taken an intermediate position, holding that prior to 1686 Leibniz held God's choice to be necessary, but that after that date—which is generally agreed on as the date by which Leibniz'

mature philosophy had taken form—he regarded God's choice as contingent (p. 69).

One thing which emerges clearly from this is that in different places Leibniz expressed contradictory opinions. Both the necessity and the contingency of God's choice can be supported textually. What is not clear is that the texts supporting necessity can be separated neatly from those supporting contingency on the basis of their date. Some of the texts Parkinson relies on are dated as late as 1698, whereas the one I had emphasized is dated (tentatively, it is true) 1680–1682. Leibniz would appear to have vacillated on this issue. It is worth pointing out, however, that the line of defense which Parkinson stresses is not a very satisfactory one. It is an uncontroversial truth of modal logic that if  $p$  is necessary and  $p$  entails  $q$ , then  $q$  is necessary. So if it is (absolutely) necessary that God choose the best, and if the existence of the best world is (hypothetically) necessary in relation to his choice, then it is (absolutely) necessary that the best world exist. The alternative view, according to which God's choice is contingent, may be unattractive in its own way, but I do not think it is incoherent.

Another topic which has been the focus of a great deal of attention lately is Leibniz' doctrine of relations. Russell had been sharply critical of Leibniz on this point,<sup>2</sup> though he was hardly precise or consistent in defining the doctrine he wished to attack, sometimes ascribing to Leibniz the view that putatively relational propositions were “no propositions at all” or were “meaningless,” and sometimes ascribing to him the view that they were reducible to non-relational propositions. Both Parkinson (20) and Rescher (21) reject the former, more extreme view, but agree with Russell in attributing the latter, weaker thesis to Leibniz. And they also agree with Russell in holding that even the weaker thesis can easily be refuted by reference to asymmetric relations. A similar interpretation is offered by Mates (19).

The most recent studies of this topic, however, suggest that Leibniz did not hold even the weaker thesis. Thus both Hintikka (15) and Ishiguro (16, 17) argue that, insofar as Leibniz did have a reductionist program, it seems to have involved the reduction of relational propositions which refer explicitly to more than one individual to propositions which ascribe complex, implicitly relational predicates to one subject. The implicitly relational

predicates would not involve a specific reference to other individuals, though they would “conceal” bound variables. So, for example, “Paris loves Helen” would be rendered by some such paraphrase as “Paris loves and, *eo ipso*, Helen is loved.” The predicates used in the paraphrase are implicitly relational because “Paris loves” is to be regarded as short for  $(\exists x)$  (Paris loves  $x$ ).

This interpretation does seem easier to reconcile with Leibniz' various attempts at the reduction of relational propositions and has the added merit of ascribing to Leibniz a program which is not obviously impossible to carry out. No doubt the connective “and *eo ipso*” is not truth-functional, but that would hardly be an objection which Leibniz would regard as a serious one. What is more difficult is to see the motivation for what Ishiguro calls “the re-writing project.”

Other areas in which familiar interpretations of Leibniz are being challenged involve such topics as “Leibniz' Law” and his attitude toward singular propositions referring to non-actual possibles. E.g., Mates (19) has called attention to passages in which Leibniz showed himself to be aware that the principle of substitutivity *salva veritate* (*eadem sunt quorum unum alteri substitui potest, salva veritate*) requires qualification in certain contexts. Mates has also suggested that the substitutivity principle might be intended, not as a criterion of identity for individuals, but as a criterion of identity for concepts. Mates does not argue for this himself, but it has subsequently been argued for by Feldman (13) and Ishiguro (17). I find Miss Ishiguro's arguments more persuasive than Feldman's, but it does seem an awkward feature of her interpretation that concept-identity does not entail the synonymy of the words that express the concept. So the concepts triangle and trilateral are the same, even though the words “triangle” and “trilateral” do not mean the same. And yet Leibniz says that by a concept he means what is signified by a name.

Perhaps more startling is another conclusion which Mates (19) does explicitly argue for. In attempting to describe Leibniz' conceptual framework and to incorporate his interpretation in the semantics of a formalized language, Mates ascribes to Leibniz “a decision to regard as false every atomic sentence that contains a non-denoting name.” As evidence Mates cites the dictum, “Nothing has no properties,” which he interprets as committing

<sup>2</sup> Couturat however, had seen Leibniz as a precursor of De Morgan, Peirce and Schröder in the development of a logic of relations. Cf. *La Logique de Leibniz*, p. 303 n.

Leibniz to this policy. But surely Ishiguro is right (17) to attempt to reinstate the more usual interpretation of this dictum. We might note that Descartes, too, subscribes to the maxim that nothing has no properties. But as Kenny has recently emphasized, it is a key doctrine of Descartes that we can make true predications about the non-existent.

*The Australian National University*

The above is hardly an exhaustive account of recent work on 17th century Continental philosophy. Much that is interesting has been left out for reasons of space and I am sure that I have missed much that would have been interesting. But perhaps the accompanying bibliography can make up some of the deficiencies of this survey.

*Received August 30, 1973*

## BIBLIOGRAPHY

The following is a selective bibliography for the period from 1960 on, prepared in the Department of Philosophy, Research School of Social Sciences, at the Australian National University. The preliminary searching and compilation was done principally by a research assistant, Mr. David Kipp, but I have also had assistance from Mrs. Jocelyn Harding and Miss Janette Paris.

Preference has been given to monographs, since the periodical literature is more adequately covered by the existing bibliographical aids, such as the *Philosopher's Index*, the *Répertoire Bibliographique de la Philosophie*, and the *Internationale Bibliographie der Zeitschriftenliteratur*. Reviews have been included only where both the review and the book under review seemed of more than usual interest. Fuller bibliographies of the Descartes and Spinoza literature for this period are to appear in Caton (2) and Freeman and Mandelbaum (32) respectively. A new journal, *Studia Leibnitiana*, has begun a continuous bibliography of current work on Leibniz.

Works dealing with more than one 17th century figure are listed under "miscellaneous."

Asterisks have been used to designate aids to scholarship which are of special interest.

### LIST OF ABBREVIATIONS USED

|                     |  |
|---------------------|--|
| <i>AF</i>           | Archivio di Filosofia  |
| <i>AGP</i>          | Archiv für Geschichte der Philosophie                            |
| <i>Arch. Hist.</i>  |  |
| <i>ex. Sci.</i>     | Archive for History of Exact Sciences                            |
| <i>AIHS</i>         | Archives Internationales d'Histoire des Sciences                 |
| <i>Aj</i>           | Ajatus   |
| <i>Am. J. Phys.</i> | American Journal of Physics                                      |
| <i>Ann. Sci.</i>    | Annals of Science  |
| <i>ANTW</i>         | Algemeen Nederlands tijdschrift voor Wijsbegeerte en Psychologie |
| <i>AP</i>           | Archives de Philosophie  |
| <i>APD</i>          | Archives de Philosophie du Droit                                 |
| <i>APQ</i>          | American Philosophical Quarterly                                 |
| <i>AS</i>           | L'Age de la Science  |
| <i>Austr. J.P.</i>  | Australasian Journal of Philosophy                               |
| <i>BjHS</i>         | British Journal for the History of Science                       |

|                 |  |
|-----------------|--|
| <i>CHM</i>      | Cahiers d'Histoire Mondiale                                    |
| <i>CJP</i>      | Canadian Journal of Philosophy                                 |
| <i>Dial.</i>    | Dialectica   |
| <i>DZP</i>      | Deutsche Zeitschrift für Philosophie                           |
| <i>EP</i>       | Etudes philosophiques  |
| <i>FS</i>       | French Studies   |
| <i>GCFI</i>     | Giornale critico della filosofia italiana                      |
| <i>GM</i>       | Giornale di metafisica   |
| <i>IJPR</i>     | International Journal for Philosophy of Religion               |
| <i>IPQ</i>      | International Philosophical Quarterly                          |
| <i>JHI</i>      | Journal of History of Ideas                                    |
| <i>JHP</i>      | Journal of the History of Philosophy                           |
| <i>JP</i>       | Journal of Philosophy  |
| <i>JR</i>       | Journal of Religion  |
| <i>Kin.</i>     | Kinesis  |
| <i>KSt.</i>     | Kantstudien  |
| <i>MLN</i>      | Modern Language Notes  |
| <i>Mon.</i>     | Monist   |
| <i>MS</i>       | Modern Schoolman   |
| <i>NDJFL</i>    | Notre Dame Journal of Formal Logic                             |
| <i>NS</i>       | New Scholasticism  |
| <i>PACPA</i>    | Proceedings of the American Catholic Philosophical Association |
| <i>PAS</i>      | Proceedings of the Aristotelian Society                        |
| <i>PBA</i>      | Proceedings of the British Academy                             |
| <i>PF</i>       | Philosophy Forum   |
| <i>Philos.</i>  | Philosophy   |
| <i>PN</i>       | Philosophia Naturalis  |
| <i>Pol. St.</i> | Political Studies  |
| <i>PPR</i>      | Philosophy and Phenomenological Research                       |
| <i>PR</i>       | Philosophical Review   |
| <i>PRdsch.</i>  | Philosophische Rundschau                                       |
| <i>REJ</i>      | Revue des études juives  |
| <i>RHSA</i>     | Revue d'histoire des sciences et de leurs applications         |
| <i>RM</i>       | Review of Metaphysics  |
| <i>RMM</i>      | Revue de métaphysique et de morale                             |
| <i>RP</i>       | Revue de Paris   |
| <i>RPFE</i>     | Revue philosophique de la France et de l'étranger              |
| <i>R. Pol.</i>  | Review of Politics   |

|           |   |
|-----------|---|
| RS        | Revue de synthèse                                 |
| RSPT      | Revue des sciences philosophiques et théologiques |
| RTL       | Revue de théologie et de philosophie              |
| St. Gen.  | Studium generale                                  |
| St. P.    | Studia Philosophica                               |
| St. Phil. | Studies in Philology                              |
| Syn.      | Synthese  |
| WW        | Wirkendes Wort                                    |
| ZPF       | Zeitschrift für philosophische Forschung          |

## ARNAULD

- \*ARNAULD, A. and NICOLE, P. *La logique, ou l'art de penser*. Edition critique présentée par Pierre Clair et Francois Girbal. Paris, Presses universitaires de France, 1965.

## BAYLE

- BAYLE, P. Historical and critical dictionary [selections]. Translated with introduction and notes by R. H. Popkin, with assistance of C. Brush. New York, Bobbs-Merrill, 1965.
- BRACKEN, H. M. "Bayle not a sceptic?" *JHI*, vol. 25, April 1964: 169-180.
- BRUSH, C. *Montaigne and Bayle: variation on the theme of skepticism*. The Hague, M. Nijhoff, 1966.
- \*LABROUSSE, E. *Inventaire critique de la correspondance de Pierre Bayle*, Paris, J. Vrin, 1961.
- \*——— *Pierre Bayle*. The Hague, M. Nijhoff, Vol. 1, Du pays de foix à la cité d'Erasmus, 1963. Vol. 2, Hétérodoxie et rigorisme, 1964.
- MASON, H. T. *Pierre Bayle and Voltaire*. London, Oxford University Press, 1963.
- "Pierre Bayle's religious views." *FS*, vol. 17, 1963: 205-217.
- REX, W. *Essays on Pierre Bayle and religious controversy*. The Hague, M. Nijhoff, 1965.
- SANDBERG, K. C. "Pierre Bayle's sincerity in his views on faith and reason." *St. Phil.*, vol. 61, January 1964: 74-84.

## CAMPANELLA

- BADALONI, N. *Tommaso Campanella*. Milano, Feltrinelli, 1965.
- BONANSEA, B. M. *Tommaso Campanella, renaissance pioneer of modern thought*. Washington, D.C., Catholic University of America Press, 1969.
- "The political thought of Tommaso Campanella." In Ryan, John K., ed. *Studies in Philosophy and the history of philosophy*, Washington, D.C., Catholic University of America, vol. 2, 1963: 211-248.
- CORSANO, A. *Tommaso Campanella*, Bari, Laterza, 1961.
- FEMIANO, S. *La metafisica di Tommaso Campanella*. Milano, Marzorati, 1968.
- GALLI, G. M. "Campanella difensore di Galileo." *Scientia*, vol. 104, 1969: 178-190.

## CORDEMOY

- \*CORDEMOY, G. de. *Oeuvres philosophiques*. Avec une étude biobibliographique. Edition critique présentée par P. Clair et F. Girbal. Paris, Presses Universitaires de France, 1968.

## DESCARTES

- AARSLEFF, H. "The history of linguistics and Professor Chomsky." *Language*, vol. 46, 1970: 570-585.
- ALLAIRE, E. B. "The circle of ideas and the circularity of the 'Meditations'." *Dialogue*, vol. 5, 1966: 131-153.
- ALLARD, J.-L. *Le mathématicisme de Descartes*. Ottawa, Editions de l'Université d'Ottawa, 1963.
- (1) ALSTON, W. "The ontological argument revisited." *PR*, vol. 69, 1960: 452-474. Reprinted in Doney (3).
- BARJONET-HURAU, M. *Descartes*. Paris, Editions sociales, 1964.
- BECK, L. J. *The metaphysics of Descartes. A study of the "Meditations"*. London, Oxford University Press, 1965.
- BENNETT, J. "The simplicity of the soul." *JP*, vol. 64, 1967: 648-660.
- BERTOCCI, P. A. "Descartes and Marcel on the person and his body: a critique." *PAS*, vol. 68, 1968: 207-226.
- BLACKWELL, R. J. "Descartes' laws of motion." *Isis*, vol. 57, 1966: 220-234.
- BOUWSMA, O. K. "I think I am." In *The Nature of Philosophical Inquiry*, ed. by J. Bobik. Notre Dame, University of Notre Dame Press, 1970: 237-251.
- BRACKEN, H. M. "Chomsky's variation on a theme by Descartes." *JHP*, vol. 8, 1970: 181-192.
- BROADIE, F. *An approach to Descartes' "Meditations"*. London, Athlone Press, 1970.
- BRÜNING, W. "Möglichkeiten und Grenzen des methodischen zweifels bei Descartes." *ZPF*, 1960: 536-552.
- BURKE, J. G. "Descartes on the refraction and the velocity of light." *Am. J. Phys.*, vol. 34, No. 5, 1966: 390-400.
- CARNES, R. D. "Descartes and the ontological argument." *PPR*, vol. 24, 1964: 502-511.
- CATON, H. "On the interpretation of the 'Meditations'." *Man and World*, vol. 3, 1970: 224-245. Reply: Dascal, M. "On the role of metaphysics in Descartes' thought." *Man and World*, vol. 4, 1971: 460-469.
- (2) ——— *The origin of subjectivity: an essay on Descartes*. New Haven, Yale University Press, 1973.
- "The theological import of Cartesian doubt." *IJPR*, vol. 1, 1970: 220-232.
- CHAVAGNE, A. *Descartes*. Bruxelles, Editions Labor, 1965.
- CHOMSKY, N. *Cartesian linguistics: a chapter in the history of rationalist thought*. New York, Harper and Row, 1966.
- COLLINS, J. *Descartes' philosophy of nature*. Oxford, Blackwell, 1971.
- CRONIN, T. J. *Objective being in Descartes and in Suarez*. Rome, Gregorian University Press, 1966.

- "Eternal truths in the thought of Descartes and of his adversary." *JHI*, vol. 21, 553-559.
- "Eternal truths in the thought of Suarez and of Descartes." *MS*, vol. 38, 1961: 269-288; vol. 39, 1961: 23-38.
- "Objective reality of ideas in human thought: Descartes and Suarez." In *Wisdom in depth. Essays in honor of Henri Renard, S.J.* Milwaukee, Bruce, 1966: 68-79.
- DE ANGELIS, E. *La critica del finalismo nella cultura cartesiana. Contributo per una ricerca.* Firenze, F. le Monnier, 1967.
- DENISOFF, E. *Descartes premier théoricien de la physique mathématique.* Louvain, Nauwelaerts, 1970.
- \*DESCARTES, R. *Descartes: philosophical letters.* Edited and translated by Anthony Kenny. Oxford, Oxford University Press, 1970.
- *Discourse on method, optics, geometry and meteorology.* Translated with introduction by P. J. Olscamp. Indianapolis, Bobbs-Merrill, 1965.
- \* — *Regulae ad directionem ingenii.* Texte critique établi par G. Crapulli avec la version hollandaise du XVII<sup>e</sup> siècle. The Hague, M. Nijhoff, 1966.
- \* — *Treatise on Man.* Translated and ed. by T. S. Hall. Cambridge, Mass., Harvard University Press, 1972.
- \* — *Oeuvres philosophiques.* ed. by F. Alquié. Paris, Garnier, 1963-1973, 3 vols.
- (3) DONEY, W., ed. *Descartes. A collection of critical essays.* New York, Doubleday, 1967.
- "Descartes' Conception of Perfect Knowledge." *JHP*, vol. 8, 1970: 387-403.
- EDGELEY, R. "Innate ideas." In *Knowledge and Necessity. Royal Institute of Philosophy Lectures*, ed. G. N. A. Vesey, London, Macmillan, 1970.
- ELLIS, B. "The Origin and Nature of Newton's Laws of Motion." In *Beyond the Edge of Certainty: Essays in Contemporary Science and Philosophy*, ed. Robert Colodny, comment by Hanson. Prentice-Hall, Englewood Cliffs, 1965.
- (4) FRANKFURT, H. G. *Demons, dreamers and madmen: the defense of reason in Descartes' "Meditations."* Indianapolis, Bobbs-Merrill, 1970. Review: C. Parsons, *JP*, vol. 69, 1972: 38-46.
- (5) — "Descartes' discussion of his existence in the Second Meditation." *PR*, vol. 75, 1966: 329-356.
- GEWIRTH, A. "The cartesian circle reconsidered," *JP*, vol. 67, 1970: 668-684.
- "Descartes: two disputed questions." *JP*, vol. 68, 1971: 288-296.
- GOUIER, H. *La pensée métaphysique de Descartes.* Paris, J. Vrin, 1962.
- GRIMM, R. "Cogito, ergo sum." *Theoria*, vol. 31, 1965: 159-173.
- GUNDERSON, K. "Descartes, La Mettrie, language and machines." *Philos.*, vol. 39, No. 149, 1964: 193-222.
- HALL, A. R. "Cartesian dynamics." *Arch. Hist. ex. Sci.*, vol. 1, No. 2, 1961: 172-178.
- HARRÉ, R. (ed.) *Early 17th Century Scientists.* Oxford, Pergamon Press, 1965.
- (6) HINTIKKA, J. "Cogito, ergo sum: inference or performance?" *PR*, vol. 71, 1962: 3-32; Replies, J. R. Weinberg, J. D. Carney, vol. 71, 1962: 483-496; Rejoinder, vol. 72, 1963: 487-496. Revised version of Hintikka's original article reprinted in Doney (3) and Sesonske and Fleming.
- JASPERS, K. "Descartes and philosophy." In his *Three essays: Leonardo, Descartes, Max Weber.* Translated by Ralph Manheim. New York, Harcourt, Brace and World, 1964: 59-185.
- (7) KENNY, A. "The Cartesian Circle and the Eternal Truths." *JP*, vol. 67, No. 19, 1970: 685-699. Comment on Gewirth.
- (8) — *Descartes: a study of his philosophy.* New York, Random House, 1968.
- (9) — "Descartes' Ontological Argument." In *Fact and Existence*, ed. J. Margolis. Oxford, Blackwell, 1969. Symposium paper, with comments by Malcolm, Penelhum, Williams and Sosa.
- FELDMAN, F. and LEVISON, A. "Anthony Kenny and the Cartesian Circle," *JHP*, vol. 9, 1971: 491-496. Reply by Kenny, 487-488.
- KOYRÉ, A. "Newton and Descartes." In his *Newtonian Studies*, London, Chapman and Hall, 1965.
- LAPORTE, J.-M. "Husserl's critique of Descartes." *PPR*, vol. 23, No. 3, 1963: 335-352.
- LAUDAN, L. "The clock metaphor and probabilism: the impact of Descartes on English methodological thought, 1650-1665." *Ann. Aci.*, vol. 22, No. 2, 1966: 73-104.
- LEFÈVRE, R. *La bataille du "cogito."* Paris, Presses universitaires de France, 1960.
- *La pensée existentielle de Descartes.* Paris, Bordas, 1965.
- MAGNUS, B. and WILBUR, J. B., eds. *Cartesian essays: a collection of critical studies.* The Hague, M. Nijhoff, 1969.
- MAHNKE, D. *Der Aufbau des philosophischen Wissens nach René Descartes.* München, A. Pustet, 1967.
- MORRIS, J. "Cartesian certainty." *Aust. J.P.*, vol. 47, 1969: 161-168. Reply: Schouls, P. A. "Cartesian certainty and the 'natural light'." *Aust. J.P.*, vol. 48, 1970: 116-119.
- "Descartes and Probable Knowledge." *JHP*, vol. 8, 1970: 303-312.
- MURALT, A. de. "Epoché-Malin Génie—Théologie de la toute-puissance divine. Le concept objectif sans objet, Recherche d'une structure de pensée." [Husserl, Descartes, Occam] *St. P.*, vol. 26, 1966: 159-191.
- NÁDOR, G. "L'importance de la doctrine heuristique de Descartes dans l'histoire de la science." *Dial.*, vol. 16, 1962: 25-38.
- NAKHNIKIAN, G. "The cartesian circle revisited." *APQ*, vol. 4, 1967: 251-255.
- NELSON, J. O. "In defence of Descartes: squaring a reputed circle." *Dial.*, vol. 3, 1964: 262-272. See also Frankfurt, H. G., "A reply to Mr Nelson," *Dial.*, vol. 4, 1965-66: 92-95, and Morris, J. "A plea for the French Descartes," *Dial.*, vol. 6, 1967: 236-239.

- NORTON, D. F. "Descartes on unknown faculties: an essential inconsistency." *JHP*, vol. 6, 1968: 245-256.
- PICARDI, F. *Il concetto di metafisica nel razionalismo cartesiano*. Milan, Editore Marzorati, 1971.
- POWELL, B. "Descartes' Machines." *PAS*, vol. 71, 1970-1971: 209-222.
- QUINTON, A. "Matter and space." *Mind*, vol. 73, 1964: 332-353.
- RÖD, W. *Descartes, Die innere Genesis des cartesianischen Systems*. Basel/München, Ernst Reinhardt Verlag, 1964.
- (10) — "Zum Problem des premier principe in Descartes' Metaphysik." *K. St.*, vol. 51, 1959-60: 176-195.
- "Zur Problematik der Gotteserkenntnis bei Descartes ['Le cercle cartésien']." *AGP*, vol. 43, 1961: 128-152.
- RODIS-LEWIS, G. "La domaine propre de l'homme chez les Cartésiens." *JHP*, vol. 2, 1964: 157-188.
- SALMON, E. G. "Mathematical roots of Cartesian metaphysics." *NS*, vol. 39, No. 2, 1965: 158-169.
- SCHALL, J. V. "Cartesianism and political theory." • *R. Pol.*, vol. 24, 1962: 260-282.
- SCHOLZ, H. *Mathesis Universalis, Abhandlungen zur Philosophie als strenger Wissenschaft*, Basel, 1961.
- \*SEBBA, G. *Bibliographia cartesiana: a critical guide to the Descartes literature 1800-1900*. The Hague, M. Nijhoff, 1964.
- "Some open problems in Descartes' research." *MLN*, vol. 75, 1960: 222-229.
- SESONSKE, A. and FLEMING, B. N., eds. *Meta-meditations. Studies in Descartes*. Belmont, California, Wadsworth, 1965.
- SHAPERE, D. "Descartes and Plato," *JHI*, vol. 24, October 1963: 572-576.
- THÉBAUT, M. "Descartes masqué?" *RP*, vol. 67, 1960: 136-152.
- TURBAYNE, C. M. *The myth of metaphor*. New Haven, Yale University Press, 1962.
- VUILLEMIN, J. *Mathématiques et métaphysique chez Descartes*. Paris, Presses universitaires de France, 1960.
- "Sur les propriétés formelles et matérielles de l'ordre cartésien des raisons." In *Homage to Martial Gueroult. L'histoire de la philosophie, ses problèmes, ses méthodes*. Paris, Fischbacher, 1964: 43-58.
- (11) WEINBERG, J. R. "Some reflections on Mr Hintikka's article." *PR*, vol. 71, 1962: 483-491.
- BROGLIE, L. d. *Galilée et l'aurore de la science moderne*. Paris, Université de Paris, Palais de la Découverte, 1965.
- BURSTYN, H. L. "Galileo's attempt to prove that the earth moves." *Isis*, vol. 53, June 1962: 161-185; Reply, E. J. Aiton, vol. 54, June 1963: 265-266; Rejoinder, vol. 54, Sept. 1963: 400-401; Comments, E. J. Aiton, vol. 56, Spring, 1965: 56-61; Reply, 61-63.
- CAMPANALE, D. "Esperienza e ragione nel pensiero di Galilei." *GM*, vol. 19, 1964: 753-778.
- CARBONARA, C. *La vita e il pensiero di Galileo Galilei*, Napoli, Libr. scientifica editrice, 1963.
- CASTELLI, E., ed. *Cusano e Galileo*, Padova, Cedan, 1964.
- CLAVELIN, M. *La philosophie naturelle de Galilée*. Essai sur les origines et la formation de la mécanique classique. Paris, A. Colin, 1968.
- COHEN, I. B. "The dynamics of the Galileo-Plato problem." In *IX<sup>e</sup> Congrès international d'Histoire des Sciences* (Barcelona-Madrid 1-7 September, 1959). Actes. Paris, Hermann, 1960: 187-196.
- DE SANTILLANA, G. *Fortune di Galileo*. Bari, Laterza, 1964. Biblioteca di cultura moderna, 586.
- "Galileo today." In his *Reflections on men and ideas*. Cambridge, Massachusetts, Institute of Technology, 1968: 57-70.
- DRAKE, S. "Galileo's 1604 fragment on falling bodies (Galileo gleanings)." *BJHS*, vol. 4, December 1969: 340-358.
- "Mathematics, astronomy and physics in the work of Galileo." In *Art, science and history in the Renaissance*. Edited by C. S. Singleton, Baltimore, Johns Hopkins, 1967: 305-330.
- DUBARLE, A. M. "Les principes exégétiques et théologiques de Galilée concernant la science de la nature." *RSPT*, vol. 50, 1966: 67-87.
- GALILEI, Galileo. *Galileo Galilei on motion and on mechanics*. Translated by I. E. Drabkin; introduction and notes by S. Drake. Madison, University of Wisconsin Press, 1960.
- GALLI, G. *L'idea di materia e di scienza fisica da Talete a Galileo*. Torino, G. Giappichelli-Editore, 1963.
- GEYMONAT, L. *Galileo Galilei: a biography and inquiry into his philosophy of science*. Translated with notes by S. Drake. Foreword by G. de Santillana. New York, McGraw-Hill, 1965. Review, McMullin, E. *MS*, vol. 47, 1970: 195-209.
- GILBERT, N. W. "Galileo and the school of Padua." *JHP*, vol. 1, 1963: 223-231.
- HALL, A. R. "Galileo and the science of motion." *BJHS*, vol. 2, 1965: 185-199.
- HANSON, N. R. "Galileo's discoveries in dynamics." *Science*, vol. 147, 1965: 471-478.
- HEINTEL, E. "Galilei und Kant." *PN*, vol. 9, 1965: 191-212.
- HUMPHRIES, W. C. "Galileo, falling bodies and inclined planes. An attempt at reconstructing Galileo's discovery of the law of squares." *BJHS*, vol. 3, No. 11, 1967: 225-244.

## GALILEO



- KAPLON, M. (ed.). *Homage to Galileo*. Cambridge, Mass., M.I.T. Press, 1965.
- LANGFORD, J. J. *Galileo, science and the church*. Foreword by S. Drake. New York, Desclée, 1966.
- LOSEE, J. "Drake, Galileo and the law of inertia." *Am. J. Phys.*, vol. 34, No. 5, 1966: 430-342.
- MORPURGO TAGLIABUE, G. *I processi di Galileo e l'epistemologia*. Milano, Edizioni di Comunità, 1963.
- NAMER, E. "Le copernicanisme expérimental de Galilée." *RPFE*, vol. 155, 1965: 217-228, 359-378.
- SALMON, E. "Galileo: physics and philosophy." *IPQ*, vol. 2, 1962: 621-628.
- SHEA, W. R. J. "Galileo's claim to fame: the proof that the earth moves from the evidence of the tides." *BJHS*, vol. 5, 1970: 111-127.

## GASSENDI

- \*GASSENDI, P. *Disquisitio metaphysica seu dubitationes et instantiae adversus Renati cartesii metaphysicam et responsum*. Texte établi, traduit et annoté par B. Rochot, Paris, J. Vrin, 1962.
- GREGORY, T. *Scetticismo ed empirismo—Studio su Gassendi*. Bari, Laterza, 1961.

## GEULINX

- DE VLEESCHAUWER, H. J. *Le problème du suicide dans la morale d'Arnold Geulincx*. Pretoria, University of South Africa, 1965.
- *Le "De virtute et primis ejus Proprietatibus" d'Arnold Geulincx et sa traduction flamande*. Van de Hoofdeuchden. Introduction et textes. Pretoria, Van Schaick, 1961.
- "Logica genuina ou le purisme logique. Kant et Geulincx." In *Kritik und Metaphysik*. Studien. Heinz Heimsoeth zum 80. Geburtstag. Hrsg. von F. Kaulbach und J. Ritter. Berlin, de Gruyter, 1966: 159-173.
- *Plans d'études au XVII<sup>e</sup> siècle, II: Le plan d'études d'Arnold Geulincx*. Pretoria, University of South Africa, 1964.
- DÜRR, K. "Arnold Geulincx und die klassische Logik des 17. Jahrhunderts." *Studium generale*, vol. 18, 1965: 520-541.
- LATTRE, A. de. *L'occasionalisme d'Arnold Geulincx, étude sur la constitution de la doctrine*. Paris, Editions de minuit, 1967.

## GROTIUS

- AUGÉ, G. "Le contrat et l'évolution du consensualisme chez Grotius." *APD*, vol. 13, 1968: 99-114.
- BULL, H. "The Grotian conception of international society." In Butterfield, H. and Wight, M., eds., *Diplomatic investigations*. Cambridge, Mass., Harvard University Press, 1966: 51-73.
- DE MICHELIS, F. *Le origini storiche e culturale del pensiero di Ugo Grozio*. Firenze, La Nuova Italia, 1967.

- DROETTO, A. *Studi groziani*. Pref. di N. Bobbio. Torino, Giappichelli, 1968.
- FIGGIS, J. N. *Political thought from Gerson to Grotius, 1414-1625*. Introduction by G. Mattingly. New York, Harper, 1960.
- GROTIUS, H. *The law of war and peace*. Books I, II, III. Translated by F. W. Kelsey, with introduction by J. B. Scott. New York, Liberal Arts Press, [s.d.].
- TOOKE, J. D. *The just war in Aquinas and Grotius*. London, Society for Promoting Christian Knowledge, 1965.

## KEPLER

- ARMITAGE, A. *John Kepler*. London, Faber and Faber, 1966.
- GERLACH, W. and LIST, M. *Johannes Kepler, Leben und Werk*. München, R. Piper, 1966.
- HANSON, N. R. "Copernican disturbance and the Keplerian revolution." *JHI*, vol. 22, 1961: 169-184.
- HOLTON, G. *Johannes Kepler et les origines philosophiques de la physique moderne*. Paris, Palais de la découverte, 1961.
- KNIGHT, D. C. *Johannes Kepler and planetary motion*. New York, F. Watts, 1962.
- KOESTLER, A. "Kepler and the psychology of discovery." In Grene, Marjorie, ed. *The logic of personal knowledge*. London, Routledge, 1961.
- *Watershed: a biography of Johannes Kepler*. New York, Doubleday, 1960.
- KOYRÉ, A. *La révolution astronomique: Copernic, Kepler, Borelli*. Paris, Hermann, 1961.
- ROSEN, E. (tr.) *Somnium*. Madison, University of Wisconsin Press, 1967.
- ROSEN, S. *The harmonious world of Johann Kepler*. Boston, Little and Brown, 1962.
- SCHNEER, C. J. "The end of Plato's problem." In his *The search for order*. New York, Harper and Row, 1960: 81-90.

## LEIBNIZ

- AARSLEFF, H. "Leibniz on Locke on Language." *APQ*, vol. 1, 1964: 165-188.
- AGASSI, J. "Leibniz's place in the history of physics." *JHI*, vol. 30, 1969: 331-344.
- ANGELELLI, I. "On identity and interchangeability in Leibniz and Frege." *NDJFL*, vol. 8, 1967: 94-100.
- ARNDT, H. W. "Die Entwicklungsstufen von Leibniz' Begriff einer Lingua Universalis." In *Achter Deutschen Kongress für Philosophie*. (Heidelberg, 1966). Das Problem der Sprache, München, Wilhelm Fink Verlag, 1967: 71-79.
- BALLARD, K. E. "Leibniz's theory of space and time." *JHI*, vol. 21, 1960: 49-65.
- BECK, L. W. *Early German Philosophy*. Cambridge, Massachusetts, Belknap Press of Harvard University Press, 1969: 196-240 devoted to Leibniz.
- BELAVAL, Y. "Le problème de l'erreur chez Leibniz." *ZPF*, vol. 20, 1966: 381-396.

- L'idée d'harmonie chez Leibniz. In *Etudes sur l'histoire de la philosophie en hommage à Martial Gueroult*. Paris, Fischbacher, 1964: 59-78.
- BROOKS, R. A. *Voltaire and Leibniz*. Genève, Droz, 1964.
- BUCHDAHL, G. "Minimum Principles in Science and Philosophy during the 17th and 18th Centuries." In *Proceedings of the tenth international congress of History of Science, Ithaca, 1962*. Herman, Paris, 1964.
- CASTELNUOVO, G. *Le origini del calcolo infinitesimale nell'era moderna*. Con scritti di Newton, Leibniz, Torricelli. Pref. di U. Forti. Milano, G. Feltrinelli, 1962.
- CIONE, E. *Leibniz*. Napoli, Libreria scientifica editrice, 1964.
- COSTABEL, P. *Leibniz et la dynamique: les textes de 1692*. Paris, Hermann, 1960.
- CUNNINGHAM, G. W. "On reason's reach, historical observations." *APQ*, vol. 6, 1969: 1-16.
- (12) CURLEY, E. M. *The root of contingency*. In Frankfurt (24).
- DE VLEESCHAUWER, H. J. *Perennis quaedam philosophia*. Pretoria, Communications of the University of South Africa, Vol. I, *Exégèse*, 1968; Vol. II, *Antécédents*, 1969.
- ERICKSON, S. A. "Leibniz on essence, existence and creation." *RM*, vol. 18, 1965: 476-487.
- (13) FELDMAN, F. "Leibniz and 'Leibniz' Law." *PR*, vol. 79, 1970: 510-522. Reply, Curley, *PR*, vol. 80, 1971: 497-501.
- (14) FRANKFURT, H. G. *Leibniz, a collection of critical essays*. Garden City, New York, Anchor, 1972.
- FURTH, M. "Monadology." *PR*, vol. 76., 1967: 169-200. Reprinted in Frankfurt (14).
- GOULD, J. A. "The concept of absolute space." *JHI*, vol. 22, 1961: 119-120.
- GRIMSLEY, R. "Kierkegaard and Leibniz." *JHI*, vol. 26, 1965: 383-396.
- HAAC, O. A. "Voltaire and Leibniz: two aspects of rationalism." *Studies on Voltaire and the 18th Century*, vol. 25, 1963: 795-809.
- HACKING, I. "The Leibniz-Carnap program for inductive logic." *JP*, vol. 68, 1971: 597-610.
- HALL, A. R. and HALL, M. B. "Clarke and Newton." *Isis*, vol. 52, 1961: 583-585.
- HEINEKAMP, A. *Das Problem des Guten bei Leibniz*. Bonn, Bouvier, 1969.
- HEINTEL, E. "Der Begriff der Erscheinung bei Leibniz." *ZPF*, vol. 20, 1966: 397-420.
- "Die beiden Labyrinth der Philosophie nach Leibniz." *PN*, vol. 10, 1967-68: 186-197.
- (15) HINTIKKA, J. "Leibniz on plenitude, relations and the 'reign of law'." *Aj*, vol. 21, 1969. Reprinted in Frankfurt (24).
- HOLZ, H. H. *Herr und Knecht bei Leibniz und Hegel*. Zur Interpretation der Klassengesellschaft. Neuwied, Luchterhand, 1968.
- HORN, J. C. *Monade und Begriff, Der Weg von Leibniz zu Hegel*. Wien-München, Oldenbourg, 1965.
- HOWE, L. T. "Leibniz on evil." *Sophia*, vol. 10, 1971: 8-17.
- HUBERTI, F. H. "Leibnizens Sprachverständnis unter besonderer Berücksichtigung des III. Buches der 'Neuen Untersuchungen über den Verstand'." *WW*, vol. 16, 1966: 361-375.
- Internationaler Leibniz-Kongress*, Hanover, 14. bis 19. November 1966. Akten. Band I: Metaphysik Monadenlehre. Wiesbaden, F. Steiner, 1968.
- Internationaler Leibniz-Kongress*, Hanover, 14. bis 19. November 1966. Akten. Band II: Mathematik, Naturwissenschaft. Wiesbaden, F. Steiner, 1969.
- Internationaler Leibniz-Kongress*, Hanover, 14. bis 19. November 1966. Akten. Band III: Erkenntnislehre, Logik, Sprachphilosophie, Editionsberichte. Wiesbaden, F. Steiner, 1969.
- (16) ISHIGURO, H. "Leibniz's theory of the ideality of relations." In Frankfurt (24).
- (17) — *Leibniz's philosophy of logic and language*. Ithaca, Cornell University Press, 1972.
- JALABERT, J. *Le Dieu de Leibniz*. Paris, Presses universitaires de France, 1960: 224.
- "Leibniz, philosophe de l'unité." *ZPF*, vol. 20, 1966: 447-457.
- JANKE, W. "Das ontologische Argument in der Frühzeit des Leibnizschen Denkens [1676-78]." *K. St.*, vol. 54, 1963: 259-288.
- *Leibniz, Die Emendation der Metaphysik*. Frankfurt/M., Verlag V. Klostermann, 1963.
- JOHNSON, A. H. "Leibniz's method and the basis of his metaphysics." *Philos.*, vol. 35, 1960: 51-61.
- KAULBACH, F. *Die Metaphysik des Raumes bei Leibniz und Kant*. Köln, Kölner Universitätsverlag, 1960.
- "Le labyrinthe du continu." *AP*, vol. 29, 1966: 507-535.
- KAUPPI, R. "Substitutivity salva veritate in Leibniz and modern logic." *Ratio*, vol. 10, 1968: 141-149.
- (18) — Über die Leibnizsche Logik: Mit besonderer Berücksichtigung des Problemes der Intension und der Extension. Helsinki, distributed by Akateeminen Kirjakauppa, 1960.
- KOYRÉ, A. and COHEN, I. B. "Case of the missing tanquam: Leibniz, Newton and Clarke." *Isis*, vol. 52, 1961: 555-566.
- "Newton and the Leibniz-Clarke correspondence with notes on Newton, Conti and Des Maizeaux." *AIHS*, vol. 15, 1962: 8-9, 63-126.
- KRÜGER, L. *Rationalismus und Entwurf einer universalen Logik bei Leibniz*. Frankfurt/M. Klostermann, 1969.
- LEIBNIZ 1646-1716. *Aspects de l'homme de l'oeuvre*. Journées Leibniz organisées au centre international de Synthèse, 28-30 mai 1966. Paris, Aubier-Montaigne, 1968.
- LEIBNIZ, G. W. *Allgemeiner politischer und historischer Briefwechsel*. Vol. VIII: 1692. Bearb. von G. Scheel u.a. Berlin, Akademie-Verlag, 1970.

- \*— *Confessio philosophi: ein Dialog*. Kritische Ausgabe mit Einleitung, Übersetzung, Kommentar von O. Saame. Frankfurt/M., V. Klostermann, 1967.
- *Confessio philosophi*. La profession de foi du philosophe. Texte, traduction et notes par Y. Belaval. Paris, J. Vrin, 1961.
- \*— *Fragmente zur Logik*. Ausgewählt, übersetzt und erläutert von Franz Schmidt. Berlin, Akademie-Verlag, 1960.
- *Kleine Schriften zur Metaphysik*. Hrsg. und übersetzt von H. H. Holz. Frankfurt/M., Insel-Verlag, 1965.
- *Le Dialogue irénique Bossuet-Leibniz, la réunion des Eglises en échec (1691-1702)* [rédigé par] F. Gaquère. Paris, Beauchesne, 1966.
- \*— and Arnauld, A. *The Leibniz-Arnauld correspondence*. Edited and translated by H. T. Mason. Introduction by G. H. R. Parkinson. Manchester, Manchester University Press, 1967.
- \*— *Logical papers* [selection] tr. with an introd. by G. H. R. Parkinson. Oxford, Clarendon Press, 1966.
- \*— *Monadology, and other philosophical essays*. Translated by P. and A. M. Schrecker. Introd. by P. Schrecker. New York, Bobbs-Merrill, 1965.
- \*— *Philosophical papers and letters* [selection]. Translated with introd. by L. E. Loemker. 2nd rev. ed. Dordrecht, Reidel, 1969.
- \*— *Philosophische Schriften*. Vol. I: 1663-1672. Bearb. von W. Kabitz und H. Schepers. Berlin, Akademie-Verlag, 1966.
- \*— *Politische Schriften*. Vol. II: 1677-1687. Bearb. von Lotte Knabe u.a. Berlin, Akademie-Verlag, 1963.
- \*— *Philosophische Schriften*. Vol. VI: *nouveaux essais*. Bearb. von A. Robinet und H. Schepers. Berlin, Akademie-Verlag, 1962.
- *Specimen quaestionum philosophicarum ex iure collectarum. De casibus perplexis. Doctrina conditionum. De legum interpretatione*. Introd. de T. Ascarelli, trad. par C. Ducouloux-Favard, préf. de A. Tunc. Paris, Dalloz, 1966.
- LENDERS, W. "The Analytic Logic of G. W. Leibniz and C. Wolff." *Syn.*, vol. 23, 1971: 147-153.
- LOEMKER, L. "Leibniz and the Herborn encyclopedists." *JHI*, vol. 22, 1961: 323-338.
- "Leibniz in our time. A survey of recent Leibniz literature." *PRdsch.*, vol. 13, 1965-66: 81-111.
- "Leibniz's conception of philosophical method." *ZPF*, vol. 20, 1966: 507-525.
- "Leibniz und die Grenzen des Empirismus." *KSt.*, vol. 56, 1966: 315-328.
- "Substance and process in Leibniz," in *Process and Divinity*, ed. W. R. Reese and Eugene Freeman, La Salle, Illinois, Open Court, 1964.
- LOMASKY, L. "Leibniz and the modal argument for God's existence." *Mon.*, vol. 54, 1970: 250-269.
- MACINTOSH, J. J. "Leibniz and Berkeley." *PAS*, vol. 71, 1970-71: 147-163.
- MCTIGHE, T. P. "Nicholas of Cusa and Leibniz's principle of indiscernibility." *MS*, vol. 42, 1964: 33-46.
- MARTIN, G. *Leibniz: Logic and metaphysics*. Translated by P. G. Lucas and K. J. Northcott. Manchester, Manchester University Press, 1964.
- (19) MATES, B. "Leibniz on possible worlds." In *Logic, Methodology and Philosophy of Science, III*, ed. by B. van Rootselaar and J. F. Staal, Amsterdam, North Holland Publishing Co., 1968: 507-529. Reprinted in Frankfurt (14).
- MOREAU, J. *Le Dieu des philosophes*. Paris, J. Vrin, 1969.
- "L'espace et les vérités éternelles chez Leibniz." *AP*, vol. 29, 1966: 483-506.
- \*MÜLLER, K., ed. *Leibniz-Bibliographie. Die Literatur über Leibniz*. Frankfurt/M., V. Klostermann, 1967.
- and KRÖNERT, G. *Leben und Werk von Gottfried Wilhelm Leibniz. Eine Chronik*. Frankfurt/M., V. Klostermann, XXIV, 1969.
- NAERT, E. *La pensée politique de Leibniz*. Paris, Presses Universitaires de France, 1964.
- *Mémoire et conscience de soi selon Leibniz*. Paris, J. Vrin, 1961.
- O'BRIANT, W. H. *Gottfried Wilhelm Leibniz's General Investigations concerning the analysis of concepts and truths*. Athens, University of Georgia Press, 1968.
- "Leibniz's preference for an intensional logic." *NDJFL*, vol. 8, 1967: 254-256.
- ODEGARD, D. "Essences and discovery: Plato, Locke and Leibniz." *Dialogue*, vol. 3, 1964: 219-234.
- (20) PARKINSON, G. H. R. *Logic and reality in Leibniz's metaphysics*. London, Oxford University Press, 1965.
- PERL, M. R. "Physics and metaphysics in Newton, Leibniz and Clarke." *JHI*, vol. 30, 1969: 507-526.
- PEURSEN, C. A. van. *Leibniz*. Translated by H. Hoskins. London, Faber and Faber, 1969.
- PHILONENKO, A. "La loi de continuité et le principe des indiscernables, étude leibnizienne." *RMM*, vol. 72, 1967: 261-286.
- POSER, H. "Zum logischen und inhaltlichen Zusammenhang der Modalbegriffe bei Leibniz." *KSt.*, vol. 60, 1969: 436-451.
- *Zur Theorie der Modalbegriffe bei G. W. Leibniz*. Steiner-Verlag, Wiesbaden, 1969.
- PRIESTLEY, F. E. L. "The Clarke-Leibniz Controversy." In *The Methodological Heritage of Newton*, ed. by R. E. Butts and J. W. Davis. Blackwell, Oxford, 1970: 34-56.
- PRIOR, A. N. "Time, existence and identity." In his *Papers on time and tense*. Oxford, Clarendon Press, 1968: 78-87.
- RESCHER, N. "Identity, substitution and modality." *RM*, vol. 14, 1960: 158-167.
- "Logical difficulties in Leibniz' metaphysics." In his *Essays in philosophical analysis*. Pittsburgh, University of Pittsburgh Press, 1969: 159-170.
- (21) — *The philosophy of Leibniz*. Englewood Cliffs, N.J., Prentice-Hall, 1967.
- Revue internationale de philosophie*, vol. 20, no. 76-77, 1966. Devoted to Leibniz. Articles by Robinet, Kneale, Costabel, Capek, et al.

- ROBINET, A. *Leibniz*. Paris, Seghers, 1962.
- ROSS, J. F. "Logically necessary existential statements." *JP*, vol. 58, 1961: 253-263.
- RUSSELL, L. J. "What is living and what is dead in the philosophy of Leibniz." *Filosofia*, vol. 19, 1968: 699-712.
- SAINATI, V. "Sulla logica leibniziana." *Filosofia*, vol. 21, 1970: 221-258.
- SAVILE, A. "Leibniz's contribution to the theory of innate ideas." *Philos.*, vol. 47, 1972: 113-124.
- SCHEPERS, H. "Möglichkeit und Kontingenz (zur Geschichte der philosophischen Terminologie von Leibniz)." *Filosofia*, vol. 14, 1963: 901-914.
- "Zum Problem der Kontingenz bei Leibniz; Die beste der möglichen Welten." In *Collegium philosophicum: Studien Joachim Ritter zum 60. Geburtstag*. Basel, Schwabe, 1965: 326-350.
- SCHMIDT, F. "Die Entwicklung der Leibnizischen Logik." *KSt.*, vol. 52, 1960-61: 43-58.
- SCHNEIDER, H-P. *Justitia universalis: Quellenstudien zur Geschichte des "christlichen Naturrechts" bei Gottfried Wilhelm Leibniz*. Frankfurt/M., V. Klostermann, 1967.
- SELLARS, W. "Méditations leibniziennes." *APQ*, vol. 2, 1965: 105-118.
- SERRES, M. *Le système de Leibniz et ses modèles mathématiques*. Paris, Presses universitaires de France, 1968: 2 vol.
- SIMONOVITS, A. *Dialektisches Denken in der Philosophie von Gottfried Wilhelm Leibniz*. Berlin, Akademie-Verlag, 1968.
- STRAUSS, M. "Die Huygens-Leibniz-Machsche Kritik in Lichte heutiger Erkenntnis." *DZP*, vol. 16, 1968: 117-120.
- TOTOK, W. and HAASE, C. (ed.). *Leibniz: sein Leben, sein Wirken, seine Welt*. Hanover, Verlag für Literatur und Zeitgeschichte, 1966.
- TYMIEŃECKA, A. T. *Leibniz' cosmological synthesis*. Assen, Van Gorcum, 1964.
- "Leibniz' metaphysics and his theory of the universal science." *IPQ*, vol. 3, 1963: 370-391.
- WIEDEBURG, P. *Der junge Leibniz, das Reich und Europa*. Teil I. Wiesbaden, Steiner, 1962.
- WILSON, M. "Leibniz and Locke on first truths." *JHI*, vol. 28, 1967: 347-366.
- "Possibility, propensity and chance: some doubts about the Hacking thesis." *JP*, vol. 68, 1971: 610-617.
- LATTRE, A. DE. "Le ferveur malebranchiste et l'ordre des raisons." *RPFE*, vol. 86, 1961: 73-86; vol. 87, 1962: 359-364; vol. 89, 1964: 90-102.
- "La volonté chez Malebranche et les contradictions de la doctrine." *RPFE*, vol. 89, 1964: 103-109.
- \*(24) MALEBRANCHE, N. *Oeuvres complètes*. 20 vols. ed. par A. Robinet, G. Lewis-Rodis, et al. Paris, J. Vrin, 1960-66.
- MERLEAU-PONTY, M. *L'Union de l'âme et du corps chez Malebranche, Maine de Biran et Bergson*. Notes prises au cours de M. Merleau-Ponty à l'Ecole Normale Supérieure (1947-1948) recueillies et rédigées par Jean Deprun, Paris, J. Vrin, 1968.
- MOREAU, J. "Saint Augustin et Malebranche." In *La Philosophie et ses problèmes. Recueil d'études offert à Monseigneur R. Jolivet*. Lyons-Paris, E. Vitte, 1960: 109-136.
- (25) NICOLASI, S. *Causalità divina e libertà umana nel pensiero di Malebranche*. Padova, Cedam, 1963.
- ROBINET, A. "Conception tragique et conception optimiste de la nature humaine dans la philosophie de Malebranche." *EP*, vol. 16, 3, 1961: 317-321.
- (26) ——— *Malebranche de l'Académie des Sciences*. Paris, Librairie philosophique, J. Vrin, 1970: 447.
- (27) ——— *Système et existence dans l'œuvre de Malebranche*. Paris, J. Vrin, 1965.
- (28) RODIS-LEWIS, G. *Nicholas Malebranche*. Paris, Presses Universitaires de France, 1963.
- (29) ROME, B. *The philosophy of Malebranche: a study of his integration of faith, reason, and experimental observation*. Chicago, Henry Regnery, 1963.
- WALTON, C. "Malebranch's ontology." *JHP*, vol. 7, 1969: 143-161.
- (30) ——— *De la recherche du bien, a study of Malebranche's science of ethics*. The Hague, M. Nijhoff, 1972.

## MERSENNE

- \*MERSENNE, M. *Correspondance du P. Marin Mersenne*. Vol. III, 2<sup>e</sup> ed., Vol. VI-IX ed. C. de Waard. Paris, Centre National de la Recherche Scientifique, 1961-69.
- *Harmonie universelle: the books on instruments*. Translated by R. Chapman. New York, W. S. Heinman, 1965.
- \*—— *Les Mécaniques de Galilée mathématicien et ingénieur du duc de Florence*. Traduites de l'italien par le P. Marin Mersenne, Edition critique par B. Rochot, Paris. Presses universitaires de France, 1966.
- ROCHOT, B. "Le P. Mersenne et les relations intellectuelles dans l'Europe du XVII<sup>e</sup> siècle." *CHM*, vol. 10, 1966-67: 55-73.

## PASCAL

- (22) BASTIDE, G., ed. *Malebranche, l'homme et l'œuvre (1638-1715)*. Paris, J. Vrin, 1967.
- BRACKEN, H. "Berkeley and Malebranche on ideas." *MS*, vol. 41, 1963: 1-15.
- (23) CONNELL, D. *The vision in God. Malebranche's scholastic sources*. Paris, Béatrice-Neuwelaerts, 1967.
- HANKINS, T. L. "Influence of Malebranche on the science of mechanics during the eighteenth century." *JHI*, vol. 28, 1967: 193-210.

- BELAVAL, Y. "Pascal savant." *RTL*, vol. 13, 1963: 1-11.
- BIRAULT, H. "Nietzsche et le pari de Pascal." *AF*, No. 3, 1962: 67-90.

- BISHOP, M. G. *Pascal: the life of genius*. New York, Greenwood Press, 1968.
- BROOME, J. H. *Pascal*. London, Edward Arnold, 1965.
- CAILLIET, E. *Pascal. The emergence of genius*. Introduction by C. S. Duthie. 2nd ed. with appendix on recent research. New York, Harper, 1961.
- CARGILE, J. "Pascal's wager." *Philos.*, vol. 41, 1966: 250-257.
- CASTELLI, E., ed. *Pascal e Nietzsche*. Padova, Cedam, 1962.
- CHAIGNE, L. *Pascal*. Paris, Wesmael-Charlier, 1962.
- ERNST, P. *Approches pascaliennes*. Pref. de J. Mesnard. Gembloux (Belg.), Duculot, 1970.
- FRIEDENTHAL, R. *Entdecker des Ich: Montaigne, Pascal, Diderot*. München, R. Piper & Co., 1969.
- "De betekenis van Pascals mathematische filosofie." *ANTW*, vol. 54, 1962: 182-193.
- GARRONE, C. G.-M. *Ce que croyait Pascal*. Paris, Mame, 1969.
- GIVONE, S. "Il problema della fede in Pascal e in James." *Filosofia*, vol. 20, 1969: 261-288.
- GOLDMANN, L. *The hidden God. A study of tragic vision in the Pensées of Pascal and the tragedies of Racine*. Translated by Philip Thody. London, Routledge and Kegan Paul, 1963.
- GOUIER, H. *Blaise Pascal, Commentaires*. Paris, J. Vrin, 1966.
- GUITION, J. *Génie de Pascal*. Paris, Aubier, 1962.
- HACKING, I. "The logic of Pascal's wager." *APQ*, vol. 9, 1972: 186-192.
- HAZELTON, R. "Pascal's wager argument." In *The heritage of Christian thought*. Essays in honor of Robert Lowry Calhoun. Edited by R. E. Cushman and E. Grislis. New York, Harper, 1965: 108-126.
- ITARD, J. "L'introduction à la géométrie de Pascal." *RHSA*, vol. 15, 1962: 269-286.
- JAMES, E. D. "The theorem of sufficient grace and the *Lettres provinciales*." *FS*, vol. 21, 1967: 205-219.
- KLEIN, Z. *La notion de dignité humaine dans la pensée de Kant et de Pascal*. Paris, J. Vrin, 1968.
- LARRAILLET, P. *Pascal et le problème de la destinée*. Paris, Nouvelles éditions latines, 1970.
- MARTINO, V. B. *Pascal nella unità del suo pensiero*. Lecce, Edizione Milella, 1962.
- MEIL, J. *Pascal and theology*. Baltimore, Johns Hopkins Press, 1970.
- MESNARD, J. *Pascal et les Roannez*. I-II. Bruges-Paris, Desclée de Brouwer, 1965.
- PAREYSON, L. *L'etica di Pascal*. Torino, G. Giappichelli, 1966.
- PENELHUM, T. "Pascal's wager." *JR*, vol. 44, 1964: 201-209.
- PLAMENATZ, J. "Pascal and Rousseau." *Pol. St.*, vol. 10, 1962: 248-263.
- SADOUN-GOUPIL, M. "L'oeuvre de Pascal et la physique moderne." *RHSA*, vol. 16, 1963: 23-52.
- SELLIER, P. *Pascal et saint Augustin*. Paris, Colin, 1970: 656.
- SINA, M. L. "Anti-Pascal" di Voltaire. Milano, Vita e Pensiero, 1970: 261.
- STEINMANN, J. *Pascal*. Translated by M. Turnell, New York, Harcourt, Brace and World, 1966: 304.
- TURNER, M. B. "Deciding for God. The Bayesian support of Pascal's wager." *PPR*, vol. 29, 1968: 84-90.

## PUFENDORF

- KRIEGER, L. *The politics of discretion: Pufendorf and the acceptance of natural law*. Chicago, University of Chicago Press, 1965.

## SPINOZA

- BECK, L. W. "Spinoza." In his *Six secular philosophers*. New York, Harper and Row, 1960: 27-41.
- BELAEF, G. *Spinoza's philosophy of law*. The Hague, Mouton, 1971.
- BOLLACHER, M. *Der junge Goethe und Spinoza*. M. Niemeyer Verlag, Tübingen, 1969.
- CHRISTIAN, W. A. "Spinoza on theology and truth." In *The heritage of Christian thought*. Essays in honor of Robert Lowry Calhoun. Edited by R. E. Cushman and E. Grislis. New York, Harper, 1965: 89-107. •
- (31) CURLEY, E. M. *Spinoza's metaphysics: an essay in interpretation*. Cambridge, Harvard University Press, 1969.
- DE DEUGD, C. "Old wine in new bottles? Tillich and Spinoza." In *Royal Institute of Philosophy. Talk of God*. New York, St. Martins, 1969: 133-151.
- *The significance of Spinoza's first kind of knowledge*. Assen van Gorcum, 1966.
- DELEUZE, G. *Spinoza*. Paris, Presses universitaires de France, 1970.
- *Spinoza et le problème de l'expression*. Paris, Editions de Minuit, 1968.
- DE LUCCA, J. "Wolfson on Spinoza's use of the *more geometrico*." *Dial.*, vol. 6, 1967: 89-102.
- EISENBERG, P. "How to understand *De Intellectus Emendatione*." *JHP*, vol. 9, 1971: 171-191.
- FEUER, L. S. *Spinoza and the rise of liberalism*. London, Mayflower, 1960.
- FRANK, D. F. *Spinoza*. Amsterdam, De Bezige Bij, 1964.
- (32) FREEMAN, E. and MANDELBAUM, M. *Spinoza: Essays in Interpretation*. La Salle, Ill., Open Court, 1973.
- CALVETTI, C. *Spinoza, I presupposti teoretici dell' irrenismo etico*. Milano, Ed. Vita e Pensiero, 1968.
- \*GIANCOTTI-BOSCHERINI, E. *Lexicon Spinozanum*. 2 vol. The Hague, Martinus Nijhoff, 1970.
- GRAM, M. S. "Spinoza, substance and predication." *Theoria*, vol. 34, 1968: 222-244.
- GRENE, M., ed., *Spinoza, a collection of critical essays*. Garden City, N.Y., Anchor, 1973.
- (33) GUEROUULT, M. *Spinoza, Vol. I: Dieu (Ethique, I)*. Paris, Editions Montaigne, 1968.
- HALLETT, H. F. *Creation, emanation and salvation: a Spinozistic study*. The Hague, M. Nijhoff, 1962.
- HAMPSHIRE, S. "Spinoza and the idea of freedom." *PBA*, vol. 46, 1960: 195-215.

- HUBBELING, H. S. *Spinoza*. Baarn, Het Wereldvenster, 1966.
- *Spinoza's methodology*. Assen, Van Gorcum, 1964.
- INQUIRY, vol. 12, No. 1, 1969. Devoted to Spinoza, with articles by Saw, Parkinson, Naess, McShea, et al.
- LACHARRIÈRE, R. *Etudes sur la théorie démocratique, Spinoza, Rousseau, Hegel, Marx*. Paris, Payot, 1963.
- LACROIX, J. *Spinoza et le problème du salut*. Paris, Presses universitaires de France, 1970.
- Les études philosophiques*. n. 3, juillet-septembre, 1972. Devoted to Spinoza, with articles by Caillois, Matheron, Moreau, Zac, et al.
- MC SHEA, R. J. *The political philosophy of Spinoza*. New York, Columbia University Press, 1968.
- MALET, A. *Le traité théologico-politique de Spinoza et la pensée biblique*. Paris, Société d'Éditions Les Belles Lettres, 1966.
- MARK, T. C. *Spinoza's theory of truth*. New York, Columbia University Press, 1972.
- MATHERON, A. *Individu et communauté chez Spinoza*. Paris, Les éditions de Minuit, 1969.
- MILLET, L. *Pour connaître la pensée de Spinoza*. Paris, Bordas, 1971.
- MISRAHI, R. *Spinoza*. Paris, Seghers, 1964.
- NAULIN, P. "La connaissance du bien et du mal selon Spinoza." *EP*, No. 3, 1969: 359-371; No. 1, 1970: 13-24.
- \*OKO, A. S., comp. *Spinoza bibliography*. Boston, Hall 1964.
- PINES, S. "Spinoza's *Tractatus theologico-politicus*, Maimonides and Kant." In *Further studies in philosophy*. Edited by O. Segal, Jerusalem, Magnes Press, 1968: 3-54.
- PRÉPOSIET, J. *Spinoza et la liberté des hommes*. Paris, Gallimard, 1967.
- REALE DILIBERTO, M. A. "L'errore come fatto teoretico nella filosofia di Spinoza." *Filosofia*, vol. 18, April 1967: 265-294.
- RÉVAH, I. S. "Aux origines de la rupture spinozienne." *REJ*, vol. 123 (s. 4, v. 3), 1964: 359-431.
- *Spinoza et Juan de Prado*. Paris, La Haye, Mouton & Co., 1959.
- RÖD, W. "Spinozas Lehre von der Societas." *Filosofia*, vol. 18, 1967: 777-806; vol. 19, 1968: 671-698.
- "Van den Hoves *Politische Waage* und die Modifikation der Hobbesschen Staatsphilosophie bei Spinoza." *JHP*, vol. 8, 1970: 29-48.
- ROUSSET, B. *La perspective finale de "l'Éthique" et le problème de la cohérence du spinozisme. L'autonomie comme salut*. Paris, J. Vrin, 1968.
- \*SPINOZA, B. *Abrégé de grammaire hébraïque* [Compendium grammaticae linguae hebraeae] Introd. trad. française et notes par J. Askinazi et J. Askinazi-Gerson. Paris, J. Vrin, 1968.
- \*— *Algebraic calculation of the rainbow, 1687*. [Facsimile of the original Dutch text]. Introd. by G. ten Doesschate. Nieuwkoop, B. de Graaf, 1963.
- STEFFEN, H. *Recht und Staat in System Spinozas*. Bonn, Bouvier, 1968.
- STRAUSS, L. *Spinoza's critique of religion*. Translated by E. M. Sinclair. New York, Schocken Books, 1965.
- TEO, K. H. "The relation of substance to attributes in Spinoza." *Kin.*, vol. 1, 1968: 15-21.
- Texte zur Geschichte des Spinozismus*. Hrsg. von Norbert Altwicker, Wissenschaftliche Buchgesellschaft, Darmstadt, 1971.
- VAN DER TAK, W. G. B. *de Spinoza*. 's-Gravenhage, Kruseman, 1961.
- WATT, A. J. "The causality of God in Spinoza's philosophy." *CJP*, vol. 2, 1972: 171-189; reply by E. Harris, *ibid.*: 191-197.
- WETLESEN, J. *A Spinoza bibliography*. Particularly on the period 1940-1967. Oslo, Universitetsforlaget, 1968. 2nd rev. ed., 1971.
- ZAC, S. *L'idée de vie dans la philosophie de Spinoza*. Paris, Presses Universitaires de France, 1963.
- *Le Morale de Spinoza*. 2<sup>e</sup> édit. revue et augmentée. Paris, Presses Universitaires de France, 1966.
- *Spinoza et l'interprétation de l'Écriture*. Paris, Presses universitaires de France, 1965.

## MISCELLANEOUS

- ALLEN, D. *Doubt's boundless sea: scepticism and faith in the Renaissance*. Baltimore, Johns Hopkins Press, 1964.
- ASHWORTH, E. J. "The doctrine of supposition in the sixteenth and seventeenth centuries." *AGP*, vol. 51, 1969: 260-285.
- "Some notes on syllogistic in the sixteenth and seventeenth centuries." *NDJFL*, vol. 11, 1970: 17-33.
- BARTLEY, W. W. "Approaches to science and scepticism." *PF*, vol. 1, 1969: 318-329.
- BEAUFRET, J. *Evidence et vérité* (Descartes et Leibniz). Paris, Vezin, 1964.
- BELAVAL, Y. *Leibniz critique de Descartes*. Paris, Gallimard, 1960.
- BIRAULT, H. "Science et métaphysique chez Descartes et chez Pascal." *AP*, vol. 27, 1964: 483-526.
- BLOCK, O.-R. "Gassendi critique de Descartes." *RPFE*, vol. 156, 1966: 217-236.
- BLONDEL, M. *Dialogues avec les philosophes, Descartes-Spinoza-Malebranche-Pascal-Saint Augustin*. Préface par H. Gouhier, Paris, Aubier, 1966.
- BOAS, G. *The happy beast in French thought of the seventeenth century*. New York, Octagon, 1966.
- BREHIER, E. *Etudes de philosophie moderne*. Paris, Presses universitaires de France, 1965.
- BUCHDAHL, G. *Metaphysics and the philosophy of science, the classical origins: Descartes to Kant*. Cambridge, Mass. M.I.T. Press, 1969.
- CLARK, J. T. "Pierre Gassendi and the physics of Galileo." *Isis*, vol. 54, 1963: 352-370.
- COHEN, I. B. *The birth of a new physics*. New York, Doubleday, 1960.

- "Quantum in se est": Newton, Kepler, Galileo, Descartes and Lucretius." *PACPA*, vol. 38, 1964: 36-46.
- CUMMINS, P. "Perceptual relativity and ideas in the mind." *PPR*, vol. 24, 1963: 202-214.
- CORSANO, A. "Campanella e Galileo." *GCFI*, vol. 44, 1965: 313-332.
- DE ANGELIS, E. *Il metodo geometrico nella filosofia del Seicento*. Pisa, Instituto di filosofia, 1964.
- "Il metodo geometrico da Cartesio a Spinoza." *GCFI*, vol. 43, 1964: 394-427.
- DE LUBAC, H. *Augustinisme et théologie moderne*. Paris, Aubier, 1965. Eng. tr. *Augustinianism and modern theology* by L. Sheppard, London, Geoffrey Chapman, 1969.
- DE VLEESCHAUWER, H. J. *More seu ordine geometrico demonstratum*. Communications of the University of South Africa, Pretoria, 1961. (Study of Descartes, Geulincx and Spinoza).
- FLECKENSTEIN, J. O. *Naturwissenschaft und Politik; Von Galilei bis Einstein*. München, Verlag Goerg D. W. Callwey, 1965.
- GOODHUE, W. W. "Pascal's theory of knowledge: a reaction to the analytical method of Descartes." *MS*, vol. 47, 1969: 15-35.
- GOUIER, H. "Les philosophes du XVII<sup>e</sup> siècle devant l'histoire de la philosophie." *Bulletin de la Société d'Etudes du XVII<sup>e</sup> Siècle*, No. 54-55, 1962: 5-16.
- GUEROUT, M. *Etudes sur Descartes, Spinoza, Malebranche et Leibniz*. Hildesheim, G. Olms, 1970.
- HEIMSOETH, H. *Atom, Seele, Monade, Historische Ursprünge und Hintergründe von Kants Antinomie der Teilung*. Mainz, Verlag der Akademie der Wissenschaften und der Literatur, Wiesbaden, 1960.
- HOROWITZ, M. C. "Pierre Charron's view of the source of wisdom." *JHP*, vol. 9, 1971: 443-457.
- JAYNES, J. "Animate motion in the seventeenth century." *JHI*, vol. 31, 1970: 219-234.
- KIRKINEN, H. *Les origines de la conception moderne de l'homme-machine; Le problème de l'âme en France à la fin du règne de Louis XIV (1670-1715)*. Helsinki, Academia Scientiarum Fennica, 1960.
- KNEALE, W. and KNEALE, M. *The development of logic*. Oxford, Clarendon Press, 1962.
- KNIGHT, D. M. "Uniformity and diversity of nature in 17th century treatises of plurality of worlds." *Organon*, No. 4, 1967: 61-68.
- KOYRÉ, A. *Metaphysics & measurement: essays in scientific revolution*. London, Chapman & Hall, 1968.
- KRAILSHEIMER, A. J. *Studies in self-interest: from Descartes to La Bruyère*. New York, Oxford University Press, 1962.
- KRISTELLER, P. O. "The myth of Renaissance atheism and the French tradition of free thought." *JHP*, vol. 6, 1968: 233-243.
- LEYDEN, W. von. *Seventeenth-century metaphysics: an examination of some main concepts and theories*. London, G. Duckworth, 1968.
- LÖWTH, K. Gott, *Mensch und Welt in der Metaphysik von Descartes bis zu Nietzsche*. Göttingen, Vandenhoeck und Ruprecht, 1967.
- MCRÆ, R. "'Idea' as a philosophical term in the seventeenth century." *JHI*, vol. 26, 1965: 175-190.
- *The problem of the unity of the sciences: Bacon to Kant*. Toronto, University of Toronto Press, 1961.
- Mélanges Alexandre Koyré: publiés à l'occasion de son soixante-dixième anniversaire*. 2 vol. Paris, Hermann, 1964.
- MIEL, J. Pascal; "Port-Royal and Cartesian linguistics." *JHI*, vol. 30, 1969: 261-271.
- NÁDOR, G. "Scientia mirabilis. Descartes et Leibniz." *Dial.*, vol. 19, 1965: 144-157.
- NORTON, D. "Leibniz and Bayle: manicheism and dialectic." *JHP*, vol. 2, 1964: 23-36.
- (34) POPKIN, R. H. *The history of scepticism from Erasmus to Descartes*. Assen, Van Gorcum, 1960; 2nd edition, revised. New York, Humanities Press, 1964.
- "Scepticism, Theology and the Scientific Revolution in the 17th Century." In *Problems in the philosophy of science*, ed. I. Lakatos and A. Musgrave, Amsterdam, North Holland, 1968; comments by Williams and Momigliano.
- "The high road to pyrrhonism." *APQ*, vol. 2, 1965: 18-32.
- (35) RESCHER, N. "The concept of non-existent possibles." In his *Essays in philosophical analysis*. Pittsburgh, University of Pittsburgh Press, 1969: 73-109.
- RISSE, W. *Die Logik der Neuzeit 2 Bd: 1640-1780*. Stuttgart-Bad Cannstatt, Friedrich Fromman Verlag, 1970.
- ROBINET, A. "Cartésianisme et leibnizianisme." *RS*, vol. 82, 1961: 73-89.
- RODIS-LEWIS, G. *Descartes et le rationalisme*. Paris, Presses Universitaires de France, 1966.
- SABRA, A. I. *Theories of light from Descartes to Newton*. London, Oldbourne, 1967.
- SEBBA, G. "What is 'History of Philosophy'?" *JHP*, vol. 8, 1970: 251-262.
- SPINK, J. S. *French free-thought from Gassendi to Voltaire*. London, Athelone, 1960.
- STANKIEWICZ, W. J. *Politics and religion in seventeenth-century France*. A study of political ideas from the Monarchomachs to Bayle, as reflected in the toleration controversy. Berkeley, University of California Press, 1960.
- TALIAFERRO, R. *The concept of matter in Descartes and Leibniz*. Notre Dame Ind., University of Notre Dame, 1964.
- VESEY, G. N. A., ed. *Reason and Reality*. Royal Institute of Philosophy Lectures, Vol. 5. London, Macmillan, 1972. Articles by Brodbeck (on Descartes), Geach and MacIntosh (on Spinoza) and Ishiguro (on Leibniz).
- VILLEY, M. "Les fondateurs de l'école de droit naturel moderne au XVII<sup>e</sup> siècle." *APD*, No. 6, 1961: 73-105.

- VUILLEMIN, J. "Sur la différence et l'identité des méthodes de la métaphysique et des mathématiques chez Descartes et Leibniz et sur la conception classique des principes de causalité et de correspondance." *AGP*, vol. 43, 1961: 267-302.
- WALKER, D. P. *The decline of hell: seventeenth-century discussions of eternal torment*. London, Routledge and Kegan Paul, 1964.
- WATSON, R. "The breakdown of Cartesian metaphysics." *JHP*, vol. 1, 1963: 177-197.
- *The downfall of Cartesianism: 1673-1712*. The Hague, M. Nijhoff, 1966.
- WHITESIDE, D. T. "Patterns of mathematical thought in the later seventeenth century." *Arch. Hist. ex. Sci.*, vol. 1, 1961: 179-388.
- YATES, F. A. *Giordano Bruno and the Hermetic tradition*. London, Routledge and Kegan Paul, 1964.



## II. THE IDENTITY OF PROPERTIES

PETER ACHINSTEIN

### I

WHAT conditions must be satisfied if properties are to be counted as identical? Before trying to answer we might note some examples of property identities that are or are like those often cited in the philosophical literature. These "paradigm cases" seem to fall into two general classes. First, there are statements involving synonymous terms or expressions, such as "the property of being a vixen = the property of being a female fox." Secondly, there are contingent statements such as "the property of having an absolute temperature of  $300^{\circ}\text{K}$  = the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule" or "blue = the color the sky had at (time)  $T$ ." If these are genuine statements of property identity, and if they are true, we should like some principles that provide necessary and sufficient conditions for their truth.

We should also like such principles to allow us to determine the truth-value of the following more problematic identity claims:

1. the property of being equilateral = the property of being equiangular
2. the property of being a simple pendulum with a length  $L$  = the property of being a simple pendulum with a period  $T = 2\pi(L/g)^{1/2}$
3. the property of being red = the property of being red and self-identical
4. the property of being the third President of the U.S. = the property of being the person who designed the University of Virginia
5. the property of having a heart = the property of having a red heart
6. the property of having a heart = the property of having kidneys

Some may claim that certain statements on this list are true (e.g., 1 and 3), and that others are false (e.g., 6). But why should we make such judgments, and what of the other examples?

It is evident that the word "property" is here being used in a very broad way. We might say that a term or expression is a property designator if (1) it purports to refer to a property (i.e., it has an appropriate syntactic form, e.g., "the property

of . . .," "—ness"); (2) it is meaningful; (3) it is ascribable to things or stuffs (the things may be concrete or abstract and the stuffs may be kinds of stuffs, e.g., copper, or particular instances, e.g., the copper in that lump); (4) it is ascribable to things or stuffs *simpliciter* and not to one thing or stuff with respect to others (i.e., we distinguish properties from relations). Some might say that "the property of being blue" or "blueness" rather than "blue" satisfies (1). But since color words are supposed to be paradigm cases of property words they will be assumed to satisfy (1).

As we shall see in Sect. VII, certain expressions do not designate the same property in each sentence in which they are used, although their sense is the same in each. Furthermore, we may wish to say that *only* things and stuffs have properties, and not events, states, processes, phenomena, or properties themselves. But there are expressions satisfying (1)–(4) that might be ascribed to events and states, etc., as well as to things and stuffs, e.g., "the property of being observed by Galileo." Accordingly, for certain expressions to determine what property if any they designate we must determine in what sentence they are being used. These proposals permit as property designators examples that would not be sanctioned by any standard non-philosophical use of the word "property" of which I am aware. However, they may not be too far removed from at least one philosophical use of this word.

### II

Several recent trends in philosophy make the issue of property identity of interest. One is the study of reduction in the sciences and the realization that one field or theory is reducible to another not simply in virtue of the fact that items are correlated via scientific laws, but in virtue of something much stronger, viz., that items, including properties, are identical. Thermodynamics, it is said, reduces to statistical mechanics in virtue of the fact (among others) that the thermodynamic property temperature is identical with the mechanical property mean molecular kinetic energy. The

mental reduces to the physical, according to one version of the psycho-physical identity theory, in virtue of the fact that mental properties are identical with physical ones. A criterion of property identity becomes important for evaluating such claims and for explicating the concept of reduction to which appeal is made. The philosophy of action is another area in which property identity is invoked. One question about acts concerns their individuation. Some philosophers define an act as the exemplification of a property by an agent at a time,<sup>1</sup> and they go on to say that acts are identical iff the agents, times, and properties are identical. They thus appeal to the concept of property identity. Ontologists see the present topic as important for clarifying the concept of a property ("no entity without identity") and also for an analysis of various ontological categories such as events, states, and processes. Kim, e.g., analyzes events as things having properties (or bearing relationships) and he holds that the event of  $x$ 's having property  $P$  and the event of  $y$ 's having property  $Q$  are identical iff either the statements " $x$  has  $P$ " and " $y$  has  $Q$ " are logically equivalent or else  $x=y$  and property  $P=$ property  $Q$ .<sup>2</sup>

In all of these cases such theorists provide either no criterion for the identity of properties or else one which they themselves admit to be inadequate. Goldman proposes that properties are identical iff "they are expressible by synonymous expressions" (p. 12), but he recognizes that this is not adequate generally since it precludes contingent property identities such as "blue=the color the sky had at  $T$ ." Kim suggests that in the latter case "at least

one of the terms of the identity refers to a property via some particular(s) that stands in a certain definite relation to it" (p. 233). This criterion does not supply a sufficient condition for a statement to be a true contingent property identity statement, since it is satisfied by "red=the color of this crow," which is false.<sup>3</sup> Nor has Kim supplied a clearly acceptable *necessary* condition. Those who speak of property identity usually want to count "red=the color of light with wavelengths 650-780  $\mu$ m" as a true contingent property identity statement, yet the light that is referred to, I think it can plausibly be argued, is not a particular or set of particulars but rather a kind (of stuff).<sup>4</sup> But even if Kim's criterion were modified to include kinds as well as particulars, and even if this were a necessary condition, we would surely want to formulate additional criteria if the concept of property identity is to be clarified.

### III

Let us begin then with some general principles of property identity that might be regarded as plausible, and see how they fare with respect to our examples. Two such principles concern explanation and confirmation.

We might hold that if property  $P=$ property  $Q$  then if an explanation  $E$  explains why something has  $P$  it ought equally to explain why it has  $Q$ , and conversely, i.e.

(A) If  $P=Q$  then  $(\forall x) (\forall E) (E \text{ explains why } x \text{ has } P \equiv E \text{ explains why } x \text{ has } Q)$ .<sup>5</sup>

An analogue of principle (A) for *events* is held to be true by Kim, who claims it to be an "unexception-

<sup>1</sup> See Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, N.J., 1970).

<sup>2</sup> Jaegwon Kim, "On the Psycho-Physical Identity Theory," *American Philosophical Quarterly*, vol. 3 (1966), p. 231. See below, Sect. XI.

<sup>3</sup> Perhaps all that Kim wants is a criterion for determining when a statement is a contingent property identity statement, not when it is a true one. Even so it would have to be amended, since as it stands it permits "New York City=the color of this crow" which is not a contingent property identity statement.

<sup>4</sup> For other counterexamples see Robert L. Causey, "Attribute—Identities in Microreductions," *The Journal of Philosophy*, vol. 69 (1972), pp. 407-422.

<sup>5</sup> Those such as Deductive-Nomological (D-N) theorists who think of explanation as a relationship between sentences could understand principle (A) as follows. Variable  $x$  is to range over singular referring expressions,  $E$  over sentences,  $P$  and  $Q$  over properties,  $\pi_1$  and  $\pi_2$  over property designators, and (A) is to be formulated as: if  $P=Q$  and  $\pi_1$  designates  $P$  and  $\pi_2$  designates  $Q$ , then  $(\forall x) (\forall E) (E \text{ explains } \langle x \text{ has } \pi_1 \rangle \equiv E \text{ explains } \langle x \text{ has } \pi_2 \rangle)$ . Others might say that explanations relate sentences to states of affairs (events, etc.) "under descriptions." For them the object of explanation might be construed as an ordered pair consisting of a state of affairs and a description, and principle (A) would be understood as: if  $P=Q$  then  $(\forall x) (\forall E) (E \text{ explains } \langle \text{the state of } x\text{'s having } P, \text{ the description } \langle x\text{'s having } P \rangle \rangle \equiv E \text{ explains } \langle \text{the state of } x\text{'s having } Q, \text{ the description } \langle x\text{'s having } Q \rangle \rangle)$ , where  $x$  ranges over things and stuffs and  $E$  over sentences.

I shall not here pursue the question of the ontology of explanans and explanandum, except to note that the following argument works where explananda are construed as sentences, as states of affairs (events, etc.) "under descriptions," and as indirect questions. If we construe the relationship between explanans and explanandum as a relationship between states of affairs *simpliciter* then the argument breaks down because we shall have to reject (3) below and say that a state of affairs can explain itself. This seems to me to be enough to show that an explanation does not relate states of affairs *simpliciter*. For more ammunition see Israel Scheffler, *The Anatomy of Inquiry* (New York, 1963), pp. 57-76 and my "The Object of Explanation."

able principle that if event  $e$  is identical with event  $e'$ , any explanans for  $e$  is an explanans for  $e'$ .<sup>6</sup> I.e., if  $e=e'$  then  $(\forall E) (E \text{ explains } e \equiv E \text{ explains } e')$ . Since Kim also claims that an event can be analyzed as a particular's having a property (or bearing a relation), he seems committed to principle (A) as well.

Perhaps we can appeal to Principle (A) in deciding how to classify some of the items on our dubious list of Sect. I. For example, we should expect that some explanation will explain why a creature has the property of having a heart but not why it has the property of having kidneys, or why it has the property of having a heart but not why it has the property of having a red heart, or why Jefferson had the property of being the third President of the U.S., but not why he had the property of being the person who designed the University of Virginia. Thus Principle (A) seems to exclude items 4, 5, and 6 on the dubious list from the class of true property identity statements.

However, there is a major problem with (A), if we accept as true certain theoretical identities and certain assumptions that scientists and most philosophers of science accept. Suppose we claim that

- (1) the property of having an absolute temperature of  $300^\circ\text{K}$  = the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule.

Scientists who defend statistical mechanics believe that it is capable of explaining thermodynamic phenomena. They say that facts about a gas's thermodynamic properties such as temperature, pressure, volume, and entropy can be explained by appeal to facts about mechanical properties of a gas's molecules such as mass, momentum, and kinetic energy. In particular, they would assent to a statement such as this:

- (2) why this (sample of) gas has the property of having an absolute temperature of  $300^\circ\text{K}$  can be explained by appeal to the fact that it has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule.

Philosophers as well as scientists also seem committed to the following:

- (3) why this (sample of) gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule cannot be explained by appeal to the fact that it has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule.

But Principle (A) is incompatible with the conjunction of (1), (2), and (3). According to (A), if the property of having an absolute temperature of  $300^\circ\text{K}$  = the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, then anything that explains why a sample of gas has the former property also explains why it has the latter. But according to (2) the fact that this gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule explains why it has the property of having a temperature of  $300^\circ\text{K}$ , so by (1) and (A) it must also explain why it has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, which violates (3).

At least one of (A), (1), (2), and (3) must be abandoned. (3) would seem to be unchallengeable on any theory of explanation that has been proposed. Virtually all philosophers who discuss property identity accept (1), or some variant of it, as a paradigm case of a true contingent property identity statement. It's difficult to imagine abandoning (2) without abandoning much if not all of the idea strongly suggested by scientists themselves that a microtheory such as statistical mechanics attempts to explain why thermodynamic systems exhibit the properties they do. (2) indeed seems to be required by the standard D-N theory of explanation whether or not (1) is accepted, since a statement that a gas has the property of having a temperature of  $300^\circ\text{K}$  is deductively derivable from the set of laws constituting kinetic theory plus the initial condition statement that the gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, and it is not derivable from the initial condition statement alone. Nor is there any reason to suppose that a defender of a contextualist approach to explanation would want to abandon (2) even if he accepted (1).<sup>7</sup> He will say that in certain contexts it is legitimate to explain why an item has a certain property by redescribing that same property in a different way, perhaps at a different level

<sup>6</sup> Kim, "Events and Their Descriptions," *Essays in Honor of Carl G. Hempel*, ed. by Nicholas Rescher (Dordrecht, 1969), p. 201.

<sup>7</sup> See Michael Scriven, "Explanations, Predictions, & Laws," *Minnesota Studies in the Philosophy of Science*, ed. by Herbert Feigl and G. Maxwell (Minneapolis, 1962), Vol. III, pp. 170-230; Sylvain Bromberger, "An Approach to Explanation," *Analytical Philosophy*, ed. by R. J. Butler (Oxford, 1965), II, pp. 72-105; Peter Achinstein, *Law and Explanation* (Oxford, 1971), ch. 4.

and attributing it to that item. If we want to uphold (1), (2), and (3) then Principle (A) must be abandoned, and with it one possible criterion of property identity.

#### IV

Let's turn next to confirmation. We might hold that if property  $P$  = property  $Q$ , then something that confirms the proposition that an item has  $P$  confirms the proposition that it has  $Q$ , and conversely. More generally, we might say that if  $P=Q$  then any evidence  $e$  which confirms the hypothesis that something has  $P$  confirms to the same degree the hypothesis that it has  $Q$ :

- (B) If  $P=Q$  then  $(\forall x)(\forall e)[c(\text{the hypothesis that } x \text{ has } P, e) = c(\text{the hypothesis that } x \text{ has } Q, e)]$ .

As in the case of explanation there are various theories of confirmation but no matter which we accept we might expect Principle (B) to settle the issue of which items to exclude from our dubious list of Sect. I. For example, we would expect that there would be evidence  $e$  that confirms that Jefferson had the property of being the third President of the U.S. but that does not equally confirm that he had the property of being the person who designed the University of Virginia, so that item 4 should be dropped.

However, as with explanation there are problems besetting the basic principle involved. Let's return to the following property identity:

- (1) the property of having an absolute temperature of  $300^\circ\text{K}$  = the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule.

Suppose we have observed that a certain sample of gas has the property of having an absolute temperature of  $300^\circ\text{K}$ , and we want to determine the degree to which this information confirms the proposition that the gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule. Ordinarily one would include in one's evidence theoretical assumptions one is making (though the usual confirmation theories, e.g. Carnap's, don't permit this, or haven't been worked out so as to make this possible). But for the point in question this is not necessary or in-

deed desirable. Let us suppose that independently of theoretical principles we want to determine the degree of confirmation of the hypothesis that this sample of gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, solely on the basis of the evidence that it has the property of having an absolute temperature of  $300^\circ\text{K}$ . Whatever one's theory of confirmation one would clearly want to claim that the degree of confirmation in this case should be less than 1 (and very likely quite small), i.e.,

- (4) the degree of confirmation of the hypothesis that this sample of gas has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, solely on the basis of the evidence that it has the property of having an absolute temperature of  $300^\circ\text{K}$ , is less than 1.

Now every confirmation theorist holds the following to be true

- (5)  $(\forall e) [c(e, e) = 1]$ .

But Principle (B) is incompatible with the conjunction of (1), (4), and (5). According to (B), if the property of having an absolute temperature of  $300^\circ\text{K}$  = the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule, then any evidence that confirms that something has the latter confirms to an equal degree that it has the former. But according to (4), the fact that a certain sample of gas has the property of having an absolute temperature of  $300^\circ\text{K}$  provides a degree of confirmation of less than 1 for the hypothesis that the sample has the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule. So, by Principle (B), if the properties are identical, it should also provide a degree of confirmation of less than 1 for the hypothesis that the sample has the property of having an absolute temperature of  $300^\circ\text{K}$ , which violates (5). Principle (B) must be abandoned if we want to retain (1), (4), and (5).<sup>8</sup>

If we reject Principle (A) for explanation and Principle (B) for confirmation then we are recognizing the following facts. Whether an explanation  $E$  explains why some individual  $a$  has a property  $P$  depends upon how  $P$  is designated. The fact that  $E$  explains why  $a$  has  $P$  when  $P$  is designated in one way does not mean that  $E$  explains why  $a$  has  $P$

<sup>8</sup> This principle also runs into difficulty with other property identity statements. Consider "red = the color John likes best." John might say of a certain dress that it has the color he likes best, and this, together with the evidence that John has generally been reliable in the past, might be said to provide a high degree of confirmation for the hypothesis that the dress has the color John likes best. But this evidence by itself would not be expected to provide equal support for the hypothesis that the dress is red.

when  $P$  is designated in every way that it can be designated.<sup>9</sup> Similarly, whether evidence  $e$  confirms the hypothesis that  $a$  has property  $P$  depends upon how  $P$  is designated. Explanation and confirmation have an essentially linguistic character. So if we are going to say that two linguistically different expressions can designate the same property then we cannot appeal to principles of explanation or confirmation as criteria for determining whether properties designated by two expressions are the same.

## V

Can principles of property identity be formulated that will avoid the previous difficulties? One principle accepted by all who talk about property identity is that two properties are not identical if something has one but not the other, i.e.,

I. If  $P=Q$  then  $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ .

I take it that this is acceptable on the grounds that it conforms to our intuitions about correct use of "same property." We simply would not use this expression if something has one property but not the other, and it is satisfied by the kinds of examples we regard as paradigms. Indeed, Principle I seems to be a special case of the general principle of the indiscernability of identicals, according to which if  $A=B$  then whatever is true of  $A$  is true of  $B$  and conversely. Here we take  $A$  and  $B$  to be properties and "what is true of  $A$ " to be "whatever item has  $A$ ."

Although Principle I won't eliminate any of the previous examples that are worrisome, it may force us to state some of our examples with more care. Thus, a sample of gas might have the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule but not the property of having a temperature of  $300^\circ\text{K}$  because it is not in a state of thermodynamic equilibrium. Accordingly, by Principle I, we must amend the example to say something like this: (According to kinetic theory) the property of having a temperature of  $300^\circ\text{K}$  = the property of being in thermodynamic equilibrium and having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule.

A second principle, which some might say should be understood as contained in the first, provides a temporal restriction. It says that if property  $P=$

property  $Q$  then something has  $P$  at or during a time  $t$  iff it has  $Q$  at or during  $t$ :

II. If  $P=Q$ , then  $(\forall x)(\forall t)(x \text{ has } P \text{ at } t \equiv x \text{ has } Q \text{ at } t)$ .

To use this principle particular attention must be given to tenses of verbs in property designators. From 1801 to 1809 Jefferson had the property of being engaged in being the third President of the U.S. Before that time he had the property of being at a future time the third President of the U.S. According to Principle II these properties are different since the times at or during which they were possessed by Jefferson were different.

There is a problem raised by this way of speaking. Consider the property of having been the third President of the U.S., and the property of being at a future time the third President of the U.S. Jefferson had the former in 1810 and the latter in 1800 (among other times). But did he have the former after he died and the latter before he was born? One temptation is to answer "No" to both questions, since there was no Jefferson to have either of these properties at those times. On this view we should distinguish the question of when Jefferson had a property  $P$  from the question of when it is correct to say that Jefferson had  $P$ . In 1972 it was correct to say that Jefferson had the property of having been the third President of the U.S. It doesn't follow that Jefferson had that property in 1972. For Jefferson to have a property at  $t$ , on this view, Jefferson must exist at  $t$ . It follows from such a view, which I shall call the "existence" view, that in 1972 Jefferson did not have the property of having been honored by a memorial in Washington, D.C. Indeed, it follows that Jefferson never had such a property (though during his lifetime he did have the property of being honored in the future by a memorial). Those who find such a consequence unacceptable might deny that for Jefferson to have a property at  $t$  he must exist at  $t$ . Such persons would not distinguish the question of when Jefferson had a property  $P$  from the question of when it is correct to say that Jefferson had  $P$ , but would claim that to answer the former question we simply answer the latter. This view will be called the "no-existence" view. There are examples on our dubious list that Principle II will exclude no matter what stand we take on this issue. Consider: the property of being the third President

<sup>9</sup> Accordingly, we must reject the following principle, an analogue of which Kim accepts for events:  $E$  is an explanation of why  $a$  has property  $P$  iff there is some term  $\pi$  which designates  $P$  and an expression  $\alpha$  which designates  $a$ , and  $E$  explains why " $\pi\alpha$ " is true.

of the U.S. = the property of being the person who designed the University of Virginia. This is ambiguous because tenses are not indicated, but let us understand it to mean: the property of having been the third President of the U.S. = the property of having been the person who designed the University of Virginia. 1810 was a time at which Jefferson had the former property but not the latter, on both the existence and no-existence views. Applying Principle II we conclude that these properties are not identical. This principle, then, if not understood as contained in Principle I, can be used to exclude examples not excluded by the latter.<sup>10</sup>

## VI

Suppose that of all those persons who have lived and will live there are exactly three with an I.Q. greater than 200 and at a time  $T$  they happen to be alone together in the Crazy Horse Saloon (CHS). It is then true that  $\forall(x)(x \text{ has the property of being a person in the CHS at } T \equiv x \text{ has the property of having an I.Q. greater than 200})$ , so Principle I is satisfied.<sup>11</sup> However, we would not conclude that the property of being a person in the CHS at  $T$  = the property of having an I.Q. greater than 200, because although it is true that anyone that has the one property in fact has the other, someone could have the one property without having (had) the other. (For similar reasons items 4 and 6 on the dubious list of Sect. I might be rejected.) More generally we might say that if property  $P$  = property  $Q$  then it is not simply true that anything that has the one has the other, it is necessarily true:

IV. If  $P=Q$  then  $N(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ .

For those properties items have at or during a time there is the stronger principle: If  $P=Q$  then  $N(\forall x)(\forall t)(x \text{ has } P \text{ at } t \equiv x \text{ has } Q \text{ at } t)$ .

The concept of necessity needed here is *de re* necessity. Principle IV commits us to saying that if  $P=Q$  then the proposition that anything with one property has the other is a necessary proposi-

tion, no matter how these properties are designated. It is really to be understood as a universal quantification, i.e., as

( $\alpha$ )  $(\forall P)(\forall Q)$  [if  $P=Q$ , then  $N(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ ],

and since the necessity operator appears within the scope of the quantifiers we have *de re* necessity. To obtain *de dicto* necessity we would need to write

( $\beta$ )  $N(\forall P)(\forall Q)$  [if  $P=Q$  then  $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ ].

$\beta$  is undoubtedly true. Indeed, if " $(\forall P)(\forall Q)$  [if  $P=Q$  then  $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ ]" is a special case of the principle of the indiscernability of identicals, then since the latter is presumably necessary  $\beta$  is true. But  $\beta$  although true will not suffice to exclude examples such as "the property of being a person in the CHS at  $T$  = the property of having an I.Q. greater than 200." ("If the property of being a person in the CHS at  $T$  = the property of having an I.Q. greater than 200 then  $(x)(x \text{ has the property of being a person in the CHS at } T \equiv x \text{ has the property of having an I.Q. greater than 200})$ " is true and necessary.) To exclude the unwanted identity statement the necessity operator must appear within the scope of the property quantifiers, thus making the necessity *de re*.

Note that this is *de re* necessity for *properties* and not for *things*. We are not committed to saying that things necessarily have certain properties but only that certain properties (if they are identical) necessarily satisfy the condition of being such that anything that has the one has the other. More precisely, let us say that a sentence is *essentialist* if it is of the form

$(\exists x_1) \dots (\exists x_n) NF \ \& \ (\exists x_1) \dots (\exists x_n) - NF$

where  $F$  is a formula whose only free variables are  $x_1, \dots, x_n$ .<sup>12</sup> Now if we believe that there are properties, that some of them are identical and satisfy Principle IV, and that some propositions of the form " $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ " are true but

<sup>10</sup> The argument is not altered if we rewrite property designators to include an indefinite date, e.g., as "the property of having been the third President of the U.S. at some time = the property of having been the person who designed the University of Virginia at some time." Again 1810 was a time at which Jefferson had the former property but not the latter.

A third principle we might consider is a spatial analogue of II: If property  $P$  = property  $Q$  then any place where  $x$  is when  $x$  has  $P$  is a place where  $x$  is when  $x$  has  $Q$ , and conversely. To say that  $p$  is a place where  $x$  is when  $x$  has  $P$  is to say that there is a time  $t$  at or during which  $x$  has  $P$  and  $x$  is at place  $p$  at or during  $t$ . Accordingly, our spatial principle would be:

II'. If  $P=Q$  then  $(\forall x)(\forall p)(\forall t)(x \text{ has } P \text{ at } t \ \& \ x \text{ is at place } p \text{ at } t \equiv x \text{ has } Q \text{ at } t \ \& \ x \text{ is at place } p \text{ at } t)$ .

However, II' is redundant if we have II, since II entails II' (but not conversely).

<sup>11</sup> Principle II is also, although I won't argue for this here.

<sup>12</sup> See Terence Parsons, "Essentialism and Quantified Modal Logic," *The Philosophical Review*, vol. 78 (1969), pp. 35-52, who uses a more complicated version to avoid certain difficulties.

not necessary, then we are committed to the following essentialist sentence:

$$(\exists P)(\exists Q)N(\forall x)(x \text{ has } P \equiv x \text{ has } Q) \ \& \ (\exists P)(\exists Q) - N(\forall x)(x \text{ has } P \equiv x \text{ has } Q).$$

If we let " $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$ " be designated by  $F$ , then we are committed to saying that some properties satisfy  $F$  necessarily and some do not. However, we are not committed to saying that there are things which have some properties essentially and other things which have those properties contingently. In short, IV commits us to essentialism for properties but not to essentialism for things.

However, there may be a commitment to the latter lurking in the background depending upon how the truth or falsity of necessity claims of the needed kind is to be established. Why do we believe that

- (1)  $(\forall x)(x \text{ has the property of being in the CHS at } T \equiv x \text{ has the property of having an I.Q. greater than 200})$

is not necessary in the sense required for property identity? We do so because we believe that there could have been someone in the CHS at  $T$  whose I.Q. was not greater than 200 and there could have been someone not in the CHS at  $T$  whose I.Q. was greater than 200. This is tantamount to believing that certain counterfactual conditionals are true. Whether *de re* necessity for things as well as properties is presupposed depends on the kinds of counterfactuals whose truth is required. To establish the necessity of (1) let us suppose we appeal to a counterfactual such as

- (2) if Mark Twain had been in the CHS at  $T$  he would have had an I.Q. greater than 200.

It seems most natural to understand this in such a way that if Mark Twain = Samuel Clemens, then if (2) is true so is

- (3) if Samuel Clemens had been in the CHS at  $T$  he would have had an I.Q. greater than 200.

If (2) is so understood then when we invoke counterfactuals to establish the necessity of (1) we are in effect committed to *de re* necessity for things. We are committed to holding that if Mark Twain, no matter how referred to, had been in the CHS at

$T$ , then Mark Twain, no matter how referred to, would have had an I.Q. greater than 200. If so principle IV will need to be replaced by

- V. If  $P=Q$  then  $(\forall x)N(x \text{ has } P \equiv x \text{ has } Q)$ ,

in which the necessity operator appears within the scope of the universal quantifier.

An alternative is to appeal to indefinite conditionals (which may not express counterfactuals) of the form:

- (4) If someone had been in the CHS at  $T$  he would have had an I.Q. greater than 200,

where this does not entail (2) or (3). We can then say that Principle IV requires the truth of (4) in order that the properties in question be identical, while Principle V requires the truth of (2) and (3). The latter principle requires the truth of conditionals about actual items in our universe, the former does not. Since on linguistic grounds the inference from a property identity statement to statements like (2) and (3) seems no less plausible than an inference to (4), since it is puzzling how to construe (4) so that it does not entail (2) and (3), and finally since there seems to be no special motivation for allowing *de re* necessity for properties but not for things, I shall adopt the stronger Principle V.

For some the fact that we invoke *de re* necessity and counterfactuals will only reinforce their suspicions about property identity. They may take this as proof that this is an obscure notion. For others this will indicate what concepts will need to be appealed to (and clarified) to explicate property identity. My aim here is only to suggest that one way to avoid having to say that all co-extensive properties are identical is to invoke a principle of necessity, that the necessity must be *de re* necessity, and that a proposition is necessary in this sense only if certain counterfactuals are true.<sup>13</sup>

## VII

Even if a notion of necessity is invoked for some property identities it may be objected that the same cannot be done for all cases. For example, it seems natural to suppose that there is no necessary relationship between blue and the color the sky had at  $T$ , or between goodness and the property that

<sup>13</sup> One other alternative might briefly be mentioned. We could return to Principle I and treat the variable  $x$  as ranging over possible entities rather than actual ones. This would also eliminate examples like the one at the beginning of the present section and would avoid a commitment to *de re* necessity and counterfactuals. In return for this, however, we would be saddled with possible entities, a concept no less in need of clarification than those being eliminated.

Plato liked best. It is not a necessary truth that the sky was blue at  $T$  or that Plato liked goodness best.

To meet this objection two views should be noted concerning identity statements involving definite descriptions. These are views formulated about statements identifying things, but I believe their supporters would extend these views to property identities. The stronger position, expressed by David Wiggins, is that a speaker is using a sentence of the form " $A$  is  $B$ " to express a genuine identity only if he is using both the terms ' $A$ ' and ' $B$ ' to make a genuine reference. He explains the latter idea by saying that in a statement a speaker makes a genuine reference with an expression  $e$  if

- (i) he is in a position (and thinks he is) to accept and answer informatively the question "Which particular are you identifying by  $e$ ?" and (ii) he is in a position to answer this question without making use of the fact that his statement is true, and (iii) he knows (and thinks he knows) how to locate (directly or indirectly) the item which he means to identify by  $e$ .<sup>14</sup>

According to Wiggins, by this criterion sentences like "the evening star is the morning star" and "Hesperus is Phosphorus" are normally used to express genuine identities, but a sentence such as "Thomas Jefferson was the third President of the U.S." is not. "The Third President of the U.S." would not normally be used to make a genuine reference in the sentence in question, and in such a case Wiggins claims that the sentence is not an identity but a predication. It says of Thomas Jefferson that he was the third U.S. President.

I take it that Wiggins would want to extend this view to property identity statements as well, although condition (iii) would have to be modified if he did, since it is not clear that properties are located. Let us suppose that a suitable modification is provided (it does not matter for our purposes what this is). On this extended view a sentence such as "blue is the color the sky had at  $T$ " would not normally be used to express a genuine property identity, since someone who uttered it would not usually be referring to any particular property by the expression "the color the sky had at  $T$ ." Wiggins does grant that there are special contexts in which a sentence like "Thomas Jefferson was the third President of the U.S." could be used to express an identity, such a context being one in which the speaker is using "the third President of the U.S." to make a genuine reference. Pre-

sumably he would say the same for "blue is the color the sky had at  $T$ ."

Let us assume for the moment that Wiggins' general theory is correct. If it is then Principle V will need to be understood as making the following assertion: if the property being referred to by  $\pi_1$  = the property being referred to by  $\pi_2$  then  $(\forall x) N(x \text{ has the property being referred to by } \pi_1 \equiv x \text{ has the property being referred to by } \pi_2)$ . Let us further suppose that in a certain special context when a speaker asserts "blue is the color the sky had at  $T$ " he is making a genuine property identity statement. But then, I think, Principle V, construed as above, is satisfied. This is because by hypothesis when our speaker asserts "blue is the color the sky had at  $T$ " he is using "the color the sky had at  $T$ " to genuinely refer to some property. Now if blue is that property, then if (counterfactually) this object were blue it would have that property. That is, if this object were blue it would have the color the sky had at  $T$ . This counterfactual would be true when "the color the sky had at  $T$ " is being used to refer to a particular property which by hypothesis is identical with the property blue.

To this it might be objected that if the object were blue but the color of the sky at  $T$  had been red then the object would not have the color the sky had at  $T$ . Two points should be noted in reply. First, in the counterfactual "if this object had  $P$  it would have  $Q$ ," which must be true if " $P=Q$ " is true, the terms ' $P$ ' and ' $Q$ ' must be used to refer to the same properties they are being used to refer to in the sentence " $P=Q$ ." (An analogous requirement regarding *meaning* would be made for a criterion of synonymy of predicates.) In the counterfactual "if this object were blue it would have the color the sky had at  $T$ ," which must be true if "blue = the color the sky had at  $T$ " is true, the expression "the color the sky had at  $T$ " must be used to refer to the same property it is used to refer to in "blue = the color the sky had at  $T$ ." This means that no additional assumptions such as that the color the sky had at  $T$  is red are allowed in connection with the counterfactual that will change the reference of this expression. Secondly, there is a use of counterfactuals that allows the definite description to be employed in such a way. In accordance with this use when we say "if this object were blue it would have the color the sky had at  $T$ " we are supposing counterfactually that

<sup>14</sup> David Wiggins, "Identity Statements," *Analytical Philosophy*, ed. by R. J. Butler, Series II (Oxford, 1965), p. 45.



this object but not the sky has a different color from the one it actually has.

If we adopt the Wiggins view, then, a sentence of the form "blue is the color the sky had at  $T$ " may be used to express an identity, in which case "the color the sky had at  $T$ " is being used to refer to a particular property, and Principle V will hold. Or else (in the more usual situation) it will not be used to express an identity and Principle V is not supposed to apply to it.

By contrast there is a more traditional view of identity statements involving definite descriptions, according to which a definite description when used in a certain sentence denotes something iff there is a unique entity that satisfies the description as it is used in that sentence. On this definition, if the color the sky had at  $T$  was blue, then in the sentence "this object has the color the sky had at  $T$ " the definite description "the color the sky had at  $T$ " denotes the color blue. A definite description may denote something even if a speaker using it does not know what it denotes or even if he believes that it does not denote anything. As Donnellan emphasizes, denoting (in this sense) and referring are distinct notions, since the latter but not the former is affected by a speaker's intentions and beliefs.<sup>15</sup> Let us say that proper names denote their bearers and that a term such as "blue" denotes the property blue. Now on the more traditional view, a sentence of the form " $A$  is  $B$ " expresses an identity and is true iff both ' $A$ ' and ' $B$ ' denote the same item. "Thomas Jefferson was the third President of the U.S." expresses an identity and is true iff "Thomas Jefferson" and "the third President of the U.S." (as used in the above sentence) denote the same individual, even if the speaker is not using the latter term to *refer* (in Wiggins' sense or something like it) to any particular individual. On this view, denotation but not reference is required for identity.

If the denotation view is accepted then Principle V requires simply that if the property denoted by  $\pi_1$  = the property denoted by  $\pi_2$  then  $(\forall x)N(x$  has the property denoted by  $\pi_1 \equiv x$  has the property denoted by  $\pi_2)$ . But this principle is satisfied even when  $\pi_1$  or  $\pi_2$  are terms such as "blue" and "the color the sky had at  $T$ ." Suppose that "blue = the color the sky had at  $T$ " is a true property identity statement. Then, on the denotation view all that is required is that the term on the right denote the

same property as the term on the left. If this is satisfied, then what Principle V requires is this:  $(\forall x)N(x$  has the property denoted by the term on the left  $\equiv x$  has the property denoted by the term on the right). What sorts of counterfactuals must be true for this to be true? Well, again, it would need to be the case that if this object were blue it would have the color the sky had at  $T$ , and that is something I suggested earlier we would grant, provided that (in this case) the denotation of "the color the sky had at  $T$ " in the consequent of this counterfactual is the same as its denotation in "blue = the color the sky had at  $T$ ."

Accordingly, on either view of identity, the reference or the denotation view, Principle V holds for property identity statements like "blue = the color the sky had at  $T$ ," provided that references or denotations of definite descriptions in the consequent of  $V$  remain the same as in the antecedent. At the beginning of the paper conditions were cited for an expression to be a property designator, but these do not specify how to determine what property, if any, a given expression designates when used in a given sentence (by a speaker). The reference and denotation views each provide such a specification. On these views an expression as used in a certain sentence (by a speaker) designates some property iff the expression is being used to refer to (denotes) that property.

This discussion shows that we cannot assume that every expression that can designate (i.e., refer to or denote) a property designates the same property on each occasion of use. The expression "the color the sky had at  $T$ " may be used in two sentences in such a way that it does not designate the same property in each, even though its sense (meaning) is the same in each. In the sentence "this object has the color the sky had at  $T$ " it may designate the color the sky in fact had at  $T$ , viz., blue. In the sentence "if the color the sky had at  $T$  had been red then it would not be the case that this object, which is blue, has the color the sky had at  $T$ ," it does not designate that property, even though its meaning is the same in both sentences. (On the reference theory it is not being used to refer to any particular property; on the denotation view it denotes the property red.) Since there are expressions that do not designate the same property in all sentences in which they are used, it is necessary to understand Principle V in such a way that the

<sup>15</sup> Keith Donnellan, "Reference and Definite Descriptions," *The Philosophical Review*, vol. 75 (1966), p. 293.

properties designated by expressions in the antecedent and consequent remain unchanged.<sup>16</sup>

Now let us return to an objection mentioned earlier. If there really is a necessary relationship between blue and the color the sky had at  $T$ , i.e., if proposition

S:  $(\forall x)N(x \text{ is blue} \equiv x \text{ has the color the sky had at } T)$

is true, then the proposition "the sky was blue at  $T$ " is necessarily true, which it surely isn't.

Assume that S is true. If we could show that S entails " $N(\text{the sky was blue at } T)$ " then we would have shown that our assumption that S is true leads to an unacceptable conclusion. But " $N(\text{the sky was blue at } T)$ " is not entailed by S. S does entail " $N(\text{the sky has the color the sky had at } T \supset \text{the sky at } T \text{ was blue})$ ." But this does not entail " $N(\text{the sky was blue at } T)$ ." S might also be said to entail " $N(\text{the sky at } T \text{ had the color the sky had at } T \supset \text{the sky at } T \text{ was blue})$ ." Now the latter might be said to entail " $N(\text{the sky at } T \text{ was blue})$ " in virtue of the fact that the antecedent inside the parentheses is a tautology. But the antecedent, "the sky at  $T$  had the color the sky had at  $T$ ," properly understood, is not a tautology since it entails the non-tautological statement "the sky at  $T$  had some color." What is tautologous is the statement "the sky at  $T$  had whatever color the sky had at  $T$ ," or more clearly "if the sky had some color at  $T$  then the sky at  $T$  had that color." But this is not how the antecedent in question should be understood.<sup>17</sup>

I conclude that property identities such as "blue = the color the sky had at  $T$ " have not been shown to violate Principle V, provided the latter is properly understood. Note that to say that such items satisfy this principle is not to say that they

are themselves necessary truths. From S it does not follow that  $N(\text{blue} = \text{the color the sky had at } T)$ . Even if the "if . . . then" in Principle V were changed to "if and only if," from S only "blue = the color the sky had at  $T$ " and not " $N(\text{blue} = \text{the color the sky had at } T)$ " would follow.<sup>18</sup>

## VIII

If principles of coextensiveness and necessity were sufficient for property identity then we should have to accept as identical all properties related by laws asserting or implying that something has the one iff it has the other. (I shall speak of these properties as "law-related.") For example, we should have to accept:

the property of being an opaque object of height  $h$  = the property of being an opaque object which casts a shadow of length  $h$  when a light ray crossing its top makes an angle of  $45^\circ$  with a line to the base.

the property of being a simple pendulum with length  $L$  = the property of being a simple pendulum with a period  $T (= 2\pi(L/g)^{1/2})$

the property of being a particle that is acted on by an unbalanced external force = the property of being a particle that is accelerating

the property of being unsupported and falling  $s$  feet = the property of being unsupported and falling  $t$  seconds ( $t = (2s/g)^{1/2}$ )

Many will feel that if the previous criteria permit these examples then they are not sufficient and further principles should be sought. One principle that might be suggested is a causal one according to which if property  $P$  = property  $Q$  then something's

<sup>16</sup> Saul Kripke, "Identity and Necessity," *Identity and Individuation*, ed. by Milton Munitz (New York, 1971), calls a term that designates the same thing in all possible worlds a rigid designator and one that does not a non-rigid designator. I believe this comes to the same thing as saying that a rigid designator is a term that refers to or denotes the same thing in each (meaningful) occurrence of its use in counterfactuals and a non-rigid designator is a term that does not. Our problem arises, then, because some property designators are not rigid designators.

<sup>17</sup> The difference is between (1)  $(\exists! C) (C \text{ is a color} \ \& \ \text{the sky had } C \text{ at } T \ \& \ \text{the sky at } T \text{ had } C)$ , and (2)  $(\forall C) [(C \text{ is a color} \ \& \ \text{the sky had } C \text{ at } T) \supset \text{the sky at } T \text{ had } C]$ , as interpretations of "the sky at  $T$  had the color the sky had at  $T$ ." (2) is a tautology, but (1), the correct interpretation, is not. The reason that (1) not (2) is the correct interpretation is this. In the proposition "the sky at  $T$  had the color the sky had at  $T$ " we are concerned with the property designated by the expression "the color the sky had at  $T$ ." This is a property possessed by all and only blue things, and it is the property we want on the right side of the identity "blue = the color the sky had at  $T$ ." If (2) were the correct interpretation then the property in question would be the property of being such that if there is a color  $C$  which the sky had at  $T$  then the sky had  $C$  at  $T$ , and this property, being tautological, is possessed by all things.

<sup>18</sup> According to Kripke, *op. cit.*, if  $A$  and  $B$  are rigid designators then  $A = B$  is a necessary truth, and presumably he would say that this holds when  $A$  and  $B$  designate properties. Even if we accept such a view we are not committed to saying that "the color the sky had at  $T$ " is a rigid property designator [and hence that  $N(\text{blue} = \text{the color the sky had at } T)$ ], and I take it that Kripke would make no such claim.

having  $P$  has some cause or effect iff its having  $Q$  has the same cause or effect.<sup>19</sup>

- VI. If  $P=Q$  then  $(\forall x)(\forall e)[x$ 's having  $P$  causes (is caused by)  $e \equiv x$ 's having  $Q$  causes (is caused by)  $e]$ .<sup>20</sup>

Let us construe the notion of cause broadly to include not only cases in which we say that  $A$  is the cause of  $B$  but those in which we speak of  $A$  as being a cause of, or part of the cause of, or a causal factor in producing,  $B$ , or more generally let us speak of  $A$  as being *causally relevant* for  $B$ . If we can show that  $x$ 's having  $P$  is causally relevant for  $e$  but  $x$ 's having  $Q$  is not, then this would suffice to show that  $P \neq Q$ .

There are problems besetting Principle VI. One can be illustrated by cases involving "mental" properties. Suppose it is true that red = the color John likes best. And suppose that Mary's dress' being red (a non-mental property) is causally relevant for Bill's having the property of being angry (a mental property). If Principle VI is true, then we must also say that Mary's dress' having the color John likes best is causally relevant for Bill's having the property of being angry—something which many would want to reject (especially if Bill has never heard of John).

In defense of Principle VI, there is a reading of the sentence "Mary's dress' having the color John likes best is causally relevant for Bill's being angry" according to which it may not be inferred that Bill knows of John. On this reading the sentence might be paraphrased as "it is Mary's dress' having the color it did that was causally relevant for Bill's being angry and this color happens to be the one that John likes best." Suppose both you and I know that red is the color that John likes best, you ask me what caused Bill to be angry, and I want to answer you (and you know this) without revealing to some third person that Mary's dress is red. I might then say that Mary's dress' having the

color John likes best is causally relevant for Bill's being angry, without committing myself to the view that Bill knows of John. Indeed, the third person might realize my intention and not take me to be committing myself to any such view.

More generally, at least some causal sentences are ambiguous and may be read in a transparent or opaque way. In their transparent readings such causal sentences do not conflict with Principle VI.<sup>21</sup> The reading of causal sentences that is needed for Principle VI is the transparent one subject to the following condition:

- (i) If  $P=Q$  and  $x=y$  then  $x$ 's having  $P$  is causally relevant for  $e$  iff  $y$ 's having  $Q$  is causally relevant for  $e$ .

This condition (which is entailed by Principle VI together with an axiom of identity) is not satisfied by the concept of explanation, either on the  $D-N$  model or on contextual theories.<sup>22</sup> On both theories a gas's having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule might explain something without its being the case that its having a temperature of  $300^\circ\text{K}$  explains it, even though the properties in question are identical.

A second problem for Principle VI concerns cases involving logically equivalent property designators. Let us say that property designators like "the property of being round" and "the property of being round and red or not red" are logically equivalent. Two property designators of the form "the property of being—" are logically equivalent if what replaces the blank are logically equivalent predicates. Now if it is true that  $x$ 's having the property designated by  $\pi_1$  is causally relevant for its having the property designated by  $\pi_2$  then does it remain true when  $\pi_1$  or  $\pi_2$  are replaced by logically equivalent property designators? For example, if a body's having the property of being acted on by an unbalanced force

<sup>19</sup> A causal principle for *event* identity is proposed by Donald Davidson, "The Individuation of Events," in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht, 1969), pp. 216–234; and one for identity of *states* is proposed by William P. Alston, "Dispositions and Occurrences," *Canadian Journal of Philosophy*, vol. 1 (1971), pp. 125–154.

<sup>20</sup> "Causes" should be understood as short for "caused," "is causing," or "will cause" in case  $x$  is a thing, and as short for any of these or for the timeless sense in case  $x$  is a type of substance.

<sup>21</sup> The same problem can arise in cases other than those involving mental properties. Suppose that red = the color of Mary's dress, and that the stop-light's being red was causally relevant for the car's stopping. Then, according to Principle VI, the stop-light's being the color of Mary's dress was also causally relevant for the car's stopping, a claim which some might want to deny on the ground that Mary's dress' having a certain color had nothing to do causally with the car's stopping. Again, however, we must recognize transparent readings of causal sentences.

<sup>22</sup> On the  $D-N$  model, according to which explanations relate sentences, the proposition corresponding to (i) would be:

If " $\pi_1 = \pi_2$ " is true and " $x = y$ " is true, then " $x$  has  $\pi_1$ " explains  $S$  iff " $y$  has  $\pi_2$ " explains  $S$ , where  $\pi_1$  and  $\pi_2$  range over predicates,  $x$  and  $y$  over singular referring expressions, and  $S$  over sentences. See footnote 5.

is causally relevant for its having the property of accelerating, then is it true that

- (a) its having the property of being acted on by an unbalanced force and being red or not red is causally relevant for its having the property of accelerating

or that

- (b) its having the property of being acted on by an unbalanced force is causally relevant for its having the property of accelerating and being red or not red?

These seem to be objectionable on the grounds that in (a) a body's being red or not red is causally irrelevant for its accelerating, and in (b) its being red or not red is not something that is caused by its being acted on by an unbalanced force. To reject (a) and (b) as false, however, is to reject the idea, generally supported by those who speak of property identity, that logically equivalent property designators designate the same property, provided that we accept Principle VI.

In the present example we seem to accept that this body's having the property of being acted on by an unbalanced force and being red or not red is causally relevant for its (having the property of) accelerating iff its having the property of being acted on by an unbalanced force is causally relevant for its accelerating and its having the property of being red or not red is also causally relevant for this, or more generally,

- (ii)  $x$ 's having the property of being  $F$  and being  $G$  is causally relevant for  $e$  iff  $x$ 's having the property of being  $F$  is causally relevant for  $e$  and  $x$ 's having the property of being  $G$  is causally relevant for  $e$ .

It also seems plausible to say that this body's having the property of being red or not red is not causally relevant for its accelerating, nor is the latter causally relevant for the former. More generally, speaking of such property designators as logically necessary,

- (iii) If  $\pi$  is a logically necessary property designator then  $x$ 's having the property designated by  $\pi$  is not causally relevant for any  $e$ , nor is  $e$  causally relevant for  $x$ 's having the property designated by  $\pi$ .

But the conjunction of assumptions (ii) and (iii) with Principle VI will require us to abandon the idea that logically equivalent property designators

designate the same property. By (iii), this body's having the property of being red or not red is not causally relevant for its accelerating. Therefore, by (ii), this body's having the property of being acted on by an unbalanced force and being red or not red is not causally relevant for its accelerating. But we do want to say that this body's having the property of being acted on by an unbalanced force is causally relevant for its accelerating. Therefore, by Principle VI, we cannot say that the property of being acted on by an unbalanced force = the property of being acted on by an unbalanced force and being red or not red. These properties, although designated by logically equivalent designators, are not identical. To avoid this conclusion we must either reject (ii), or reject (iii) and say that this body's having the property of being red or not red is causally relevant for its accelerating (which is a bitter pill to have to swallow), or reject causal Principle VI.

One response would simply be to accept the idea that logically equivalent property designators do not always designate the same property. In defense of this proposal we might recall that "the color the sky had at  $T$ " does not designate the same property in each sentence in which it is used, e.g., in "blue = the color the sky had at  $T$ " and "if red were the color the sky had at  $T$  then . . ." Yet (counting a property designator as logically equivalent with itself) we have here a case of logically equivalent property designators that do not designate the same property in two sentences. On the other hand, "the color the sky had at  $T$ " is not a *rigid* property designator (see footnote 16). It might be agreed that non-rigid property designators that are logically equivalent do not always designate the same property, but it might be held that rigid ones do. And the examples considered above—"the property of being acted on by an unbalanced force and being red or not red" and "the property of being acted on by an unbalanced force"—are rigid property designators. If we want to hold that rigid logically equivalent property designators designate the same property we cannot retain the conjunction of (ii), (iii), and Principle VI.

There is the following alternative.<sup>23</sup> We could retain (iii) and Principle VI but modify (ii) by restricting the property designators "the property of being  $F$ " and "the property of being  $G$ " to those which are not logically necessary. On this proposal, although (iii) prevents us from asserting that this

<sup>23</sup> I am here indebted to Dale Gottlieb.

body's having the property of being red or not red is causally relevant for its accelerating, it does not prevent us from asserting that this body's having the property of being acted on by an unbalanced force and being red or not red is causally relevant for its accelerating. And if we do say the latter then (ii) as modified does not require us to conclude that this body's having the property of being red or not red is causally relevant for its accelerating [in violation of (iii)]. Such a proposal would permit (but not require) us to assert that rigid logically equivalent property designators designate the same property. If we adopt it, make this assertion, and retain Principle VI, then we shall have to accept propositions like (a) and (b) above as consequences. Although this may seem odd there are two mitigating considerations.

First, it may seem odd because, e.g. (a) seems to imply that this body's having the property of being red or not red is causally relevant for its accelerating. But that implication is now blocked by the proposed restriction on (ii). Secondly, since it does not follow that this body's having the property of being red or not red is causally relevant for its accelerating no new causal claim is being made by "this body's having the property of being acted on by an unbalanced force and being red or not red is causally relevant for its accelerating" over and above that made by "this body's having the property of being acted on by an unbalanced force is causally relevant for its accelerating." Reference in the former statement to being red or not red carries with it no new causal commitments, nor does it alter any such commitments already made by the latter. It is redundant. Moreover, even with the present restriction on assumption (ii) we are still not permitted to assert a statement such as "this body's having the property of being acted on by an unbalanced force and being red is causally relevant for its accelerating," if its having the property of being red is not causally relevant for its accelerating (since "the property of being red" is not a logically necessary property designator).<sup>24</sup>

An issue related to this has been raised by Dretske.<sup>25</sup> One and the same expression may be

used with different emphases and the truth-value of a causal claim can change with a change in emphasis. For example, it might be argued that although Socrates' drinking hemlock was causally relevant for his having the property of having *died* in prison it was not causally relevant for his having the property of having died *in prison*, i.e., the drinking of hemlock caused it to be the case that Socrates died but did not cause it to be the case that it was in prison where he died. One response to this might be to deny it by saying that although such a claim holds for explanation it does not hold for causation. I find little plausibility in such a proposal and suggest instead that we accept the above claim and conclude that as used in sentences of the form "his drinking hemlock was causally relevant for Socrates' having the property of \_\_\_\_" the expressions "the property of having *died* in prison" and "the property of having died *in prison*" designate different properties. (Whether we say that these expressions also have different meanings is a question I shall not pursue here.)<sup>26</sup> Accordingly, if we consider the sentences

- (1) his drinking hemlock was causally relevant for Socrates' having the property of having *died* in prison
- (2) his drinking hemlock was causally relevant for Socrates' having the property of having died *in prison*,

we can say that (1) is true and (2) is false because the property which according to (1) Socrates was caused to have is not the same as the property that according to (2) he was caused to have.

However, if we do say this then, I think, Principle VI will force us to abandon even the view that all rigid logically equivalent property designators designate the same property. The property designators appearing in (1) and (2) are, I shall assume, rigid. By usual standards the predicates "*died* in prison" and "*died in prison*" would be regarded as logically equivalent, even though they contain different emphases. A sentence which consists of a singular term followed by one predicate entails a sentence which consists of the same singular term

<sup>24</sup> It might be thought that we could avoid having to deny that rigid logically equivalent property designators designate the same property simply by disallowing an assumption we have been making so far, viz., that (what we have been calling) logically necessary property designators designate properties, while insisting that an expression of the form "the property of being *F* and being *G*" designates a property iff "the property of being *F*" and "the property of being *G*" do. But this proposal will not work, since in accordance with it "the property of being round and being red or not red" does not designate a property, and if "the property of being round" does designate a property, then we have as a consequence that logically equivalent designators do not designate the same property.

<sup>25</sup> Fred I. Dretske, "Contrastive Statements," *The Philosophical Review*, vol. 81 (1972), pp. 411-437.

<sup>26</sup> Dretske would argue that they have the same meaning; see his discussion, p. 426.

followed by the other predicate. Accordingly, by the previous definition "the property of having died in prison" and "the property of having died in prison" are logically equivalent property designators. But if sentence (1) above is true and (2) is false, then causal Principle VI requires us to conclude that these property designators, though rigid and logically equivalent, do not designate the same properties in (1) and (2). One who still wants to produce some true proposition about logically equivalent property designators and identity might try to modify the above account of logically equivalent property designators, or he might revise his thesis further by saying that two rigid property designators that are logically equivalent and that have corresponding emphases designate the same property. In the latter case what counts as "corresponding emphases" would need to be spelled out, but I shall not pursue the matter any further here. Suffice it to say that if we accept Principle VI there are reasons for thinking that the idea that all rigid logically equivalent property designators designate the same property cannot be accepted as it stands.

Turning now to a fourth assumption about the notion of cause in Principle VI, it seems reasonable to say:

- (iv) If  $x$ 's having  $P$  is causally relevant for  $e$  then  $e$  is not causally relevant for  $x$ 's having  $P$ .

Letting " $e$ " in (iv) be " $x$ 's having  $P$ " we can show that (iv) entails

- (v)  $x$ 's having  $P$  is not causally relevant for  $x$ 's having  $P$ .

And from (v) and an axiom of identity we obtain

- (vi) If  $P=Q$  then  $x$ 's having  $P$  is not causally relevant for  $x$ 's having  $Q$  nor is  $x$ 's having  $Q$  causally relevant for  $x$ 's having  $P$ .

Accordingly, the difficulty generated by the explanatory principle (A) in Sect. III cannot be generated by the causal Principle VI, since the difficulty depended on our accepting the proposition that a gas's having the property of having a temperature of  $300^\circ\text{K}$  can be explained by appeal to its having the property of having a mean molecular kinetic energy of  $6.21 \times 10^{-21}$  joule. The analogous non-explanatory causal proposition is blocked by (vi) above if the properties are identical.

Let me now formulate the causal principle as follows, where causal relevance is taken to satisfy assumptions (i)–(vi):

- VII. If  $P=Q$  then  $(\forall x)(\forall e) [x$ 's having  $P$  is causally relevant for  $e$  (or conversely)  $\equiv x$ 's

having  $Q$  is causally relevant for  $e$  (or conversely)].

Some of the assumptions about the notion of cause required for this principle are redundant. We said that (iv) entails (v), and that (v) plus an axiom of identity entails (vi). Assuming, then, that there are no other entailments among our assumptions we are left with (i)–(iv) as logically independent assumptions. And (i), though not (ii), (iii), or (iv), is entailed by Principle VII itself.

Is there a use of causal terms that satisfies these assumptions? If there is then we can invoke Principle VII without getting into difficulties of the sort generated by the earlier explanatory Principle (A). (This is guaranteed by (iv) which entails (vi).) And we can avoid having to deny that sentences such as "blue = the color the sky had at  $T$ " are, or can be used to express, true property identities [guaranteed by (i)]. As urged above, I believe that there is a use of causal terms which satisfies (i), i.e., there is a transparent reading of causal sentences. Moreover, that such a use of causal terms should satisfy (iii) and (iv) seems quite reasonable. (iii) is what we invoked to preclude such oddities as "this body's having the property of being red or not red is causally relevant for its accelerating." And (iv) is an instance of the more general principle that if  $A$  is a cause of  $B$  then  $B$  is not a cause of  $A$ . Whether there is a use of causal terms that satisfies amended but not unamended (ii) is something I find less clear. (I do think that there is such a use that satisfies amended (ii), but the question is whether it satisfies the unamended version as well, i.e., whether (ii) holds when one or more of the property designators is logically necessary.) I find it at least odd to say that this body's having the property of being acted on by an unbalanced force and being red or not red is causally relevant for its accelerating, even where from this, by amended (ii), we cannot conclude that this body's having the property of being red or not red is causally relevant for its accelerating. Whether this oddness amounts to falsity, as required by unamended (ii), or simply to redundancy, as allowed by amended (ii), I am not prepared to say.

## IX

Suppose we agree that there is a use of causal terms that satisfies the previous assumptions (whether (ii) is amended or not will not matter for the question now to be considered). Invoking the causal Principle VII how might we argue that not

all law-related properties are identical, i.e., how might examples at the beginning of the previous section be excluded?

Consider a certain object which is opaque, has a height  $h$ , and casts a shadow whose length is  $h$  when a ray of light crossing the top of the object makes an angle of  $45^\circ$  with a line to the base. This object's having the property of being opaque and its having the property of having a height  $h$  are, I think it is reasonable to say, both causally relevant for its having the property of casting a shadow whose length is  $h$  when, etc. Its being opaque causes it to cast a shadow in the first place, and its having a height  $h$  is one of the things that makes that shadow have the length it does. If so then by assumption (ii) (unamended or amended) its having the property of being opaque and having a height  $h$  is causally relevant for the effect in question. Moreover, none of the other assumptions regarding causal relevance is violated by such a claim. (This claim states nothing incompatible with the fundamental assumptions (i), (iii), or (iv).) However, as required by (v), which is entailed by (iv), this object's having the property of casting a shadow whose length is  $h$  when, etc. is not causally relevant for its having the property of casting a shadow whose length is  $h$  when, etc. But if the property of being opaque and having a height  $h$  and the property of casting a shadow whose length is  $h$  when, etc., are identical, then by Principle VII we should have to reject just such a causal claim. So VII and (v) preclude these properties from being identical.

Similar arguments can be given for at least some of the remaining examples at the beginning of Sect. VIII. Thus (at least a Newtonian would argue) an object's having the property of being a particle that is acted on by unbalanced force is causally relevant for its having the property of being a particle that is accelerating. In this case and the previous one the having of one of the properties cited is causally relevant for the having of the other, but if these properties are identical this is precluded by Principle VII and assumption (v), so the properties cannot be identical.

In both of these examples we assume that if  $x$  has  $P$  at or during time  $t$  then  $x$  has  $Q$  at or during  $t$ . Otherwise Principle II (Sect. V) would be violated and the properties in question could be shown to be non-identical without invoking the causal Principle VII. But if these properties are possessed simultaneously then the arguments above depend upon accepting the following additional assumption about causal relevance:

- (vii)  $x$ 's having  $P$  can be causally relevant for  $x$ 's having  $Q$  even if  $(\forall t)(\forall x)(x \text{ has } P \text{ at or during } t \equiv x \text{ has } Q \text{ at or during } t)$ .

Such an assumption may be rejected by those Humeans who insist on temporal priority of cause and effect. This insistence, I should claim, is by no means always justified. In particular it is not justified when the causes and effects in question are things (or types of substances) having properties, i.e., when they are what I should call *states of affairs* (or what Kim, obviously stretching the word, calls "events"). One state of affairs may be causally relevant for another even if both exist at or during the same time. Oxygen's having the property of being present at a time  $t$  can be causally relevant for the wood's having the property of burning at  $t$ . Copper's having the property of having one electron in its outer shell is causally relevant for its having the property of being a good conductor of electricity even though these states of affairs if they exist at all in time are not such that one precedes the other.

However, we need not depend on assumption (vii) to show that examples at the beginning of the previous section violate Principle VII. Consider the case in which it is claimed that the property of being unsupported and falling  $s$  feet = the property of being unsupported and falling  $t$  seconds, and suppose that  $x$  is a time bomb set to explode when it is unsupported and falls for  $t$  seconds. Its being unsupported releases a pin connected to a clock, which then starts, and when the clock runs  $t$  seconds it detonates the bomb. If these conditions were in fact satisfied and the bomb exploded then  $x$ 's having the property of being unsupported was causally relevant for its exploding, since its being unsupported released the pin and thus started the clock. Also  $x$ 's having the property of falling  $t$  seconds was causally relevant for its exploding, since the bomb was rigged to detonate when the clock in the bomb ran for  $t$  seconds. But then, by assumption (ii),  $x$ 's having the property of being unsupported and falling  $t$  seconds was causally relevant for its exploding. However, since the bomb was not rigged for distance,  $x$ 's having the property of falling  $s$  feet was not causally relevant for its exploding.

Things would be different if the bomb had been connected to an altimeter measuring distance fallen as a function of differences in atmospheric pressure. Given *this* situation  $x$ 's having the property of falling  $s$  feet could have been causally relevant for its exploding. If  $x$ 's having the property of falling



$s$  feet was not causally relevant for its exploding, then by assumption (ii) its having the property of being unsupported and falling  $s$  feet was not causally relevant for this. We have a case, then, in which  $x$ 's having the property of being unsupported and falling  $t$  seconds is causally relevant for  $x$ 's exploding but its having the property of being unsupported and falling  $s$  feet is not causally relevant for  $x$ 's exploding. Principle VII is thus violated and the properties involved are not identical.

If the examples offered above are correct they show the following to be true of causal relevance:

- (viii) It can be the case that  $x$ 's having  $P$  is causally relevant for  $e$  and  $x$ 's having  $Q$  is not causally relevant for  $e$  even though  $(\forall x)N(x \text{ has } P \equiv x \text{ has } Q)$ .

If (viii) is false then the causal Principle VII will not suffice to show that not all law-related properties are identical.

## X

In Sect. VI it was argued that by appeal to a principle of necessity we can avoid saying that if properties are coextensive they are identical. But such a principle has two potential drawbacks. It commits us to *de re* necessity, and it does not rule out enough examples, since it does not provide a way to avoid saying that all law-related properties are identical. To deal with the latter problem appeal was made to a causal principle. Now in view of difficulties associated with our principle of necessity perhaps this causal principle will suffice to rule out the very examples we want the principle of necessity to exclude. If so then we might be able to avoid reference to necessity.

The causal principle may very well do this job. For example, we might suppose that  $x$ 's having the property of being in the CHS at  $T$  was causally relevant for something (say his getting drunk) for which his having the property of having an I.Q. greater than 200 was not causally relevant. If so then Principle VII will suffice to preclude "the property of being in the CHS at  $T$  = the property of having an I.Q. greater than 200," without needing to invoke Principle V. The problem is that, as will now be shown, our causal principle does not really exclude all the examples we want excluded, and to make it suffice we shall need to incorporate *de re* necessity within it.

As we have been using the term "causal relevance," if  $x$ 's having  $P$  is causally relevant for  $e$

then  $x$  did or will have  $P$ ,  $e$  was or will be the case, and  $x$ 's having  $P$  was or will be causally relevant for  $e$ . Now suppose there never was or will be a simple pendulum with a (particular) length  $L$  or period  $T$ . If not then "the property of being a simple pendulum with length  $L$  = the property of being a simple pendulum with period  $T$ " will satisfy Principle VII. Since by hypotheses no item has either property, sentences of the form " $x$ 's having  $P$  is causally relevant for  $e$ " will be false, since for such sentences to be true it would need to be the case that  $x$  has  $P$ . But if such sentences are false then both sides of the biconditional in the consequent of VII are false and hence the biconditional is true. This means that if there never was or will be such a pendulum then Principle VII will not suffice to show that the properties in question are not identical. What seems required at this point is to say that if the properties in question are identical then if some object were a simple pendulum whose length is  $L$  and its having this property were causally relevant for  $e$  then it would be a simple pendulum whose period is  $T$  and its having that property would be causally relevant for  $e$ .

According to Principle V what is required for  $P$  and  $Q$  to be identical is not simply that  $P$  and  $Q$  be coextensive but that  $P$  and  $Q$  be necessarily coextensive. By analogy, in the case of VII, it seems, what is required for  $P$  and  $Q$  to be identical is not simply that items that actually have  $P$  and  $Q$  be such that their having  $P$  happens to have all and only those actual causes and effects that their having  $Q$  has. What is also required is that necessarily if the having of  $P$  causes or is caused by something then the having of  $Q$  causes or is caused by that same thing, i.e.,

- VIII. If  $P = Q$  then  $(\forall x)(\forall e)N[x \text{'s having } P \text{ is causally relevant for } e \text{ (or conversely)} \equiv x \text{'s having } Q \text{ is causally relevant for } e \text{ (or conversely)}]$ .

The necessity invoked here is again *de re* necessity since the necessity operator appears within the scope of quantifiers. As with Principle V, if we try to change this to *de dicto* necessity by bringing the necessity operator in front of the entire conditional we shall not be able to exclude examples like "the property of being a simple pendulum with length  $L$  = the property of being a simple pendulum with period  $T$ ," where nothing has either property. For " $N$  (if the property of being a simple pendulum with length  $L$  = the property of being a simple pendulum with period  $T$  then the consequent of



VII is satisfied)" is true whether or not anything has the properties in question.

One alternative to Principle VIII would be to abandon the idea that if  $x$ 's having  $P$  is causally relevant for  $e$  then  $x$  did or will have  $P$  and  $e$  is or will be the case. On this proposal we could say that  $x$ 's having the property of being a simple pendulum with a length  $L$  is causally relevant for its having the property of being a simple pendulum with a period  $T$  without supposing that  $x$  did or ever will have either of these properties. The claim that  $x$ 's having  $P$  is causally relevant for  $e$  would be understood as entailing "if  $x$  were to have  $P$  its having  $P$  would be causally relevant for  $e$ ." But this would not avoid the idea of necessity, since it would in effect build that idea into the concept of causal relevance. " $x$ 's having  $P$  is causally relevant for  $e$ " is now understood as entailing " $N(x$ 's having  $P$  is causally relevant for  $e$ )."

Moreover, this proposal would involve making some dubious assertions, since presumably there will be cases in which it will be true but not necessarily true that  $x$ 's having  $P$  was causally relevant for  $e$ .

Another alternative is to retain the original notion of causal relevance but abandon Principle VIII and regard the causal Principle VII as sufficient for property identity. This would entail accepting the view that all law-related properties that nothing ever has are identical, which seems unintuitive. Moreover, there seems to be something right about requiring of any  $x$  that if it were to be the case that  $x$ 's having  $P$  is causally relevant for  $e$  then it would be the case that  $x$ 's having  $Q$  is too, in order that properties be identical, just as it seems right to require of any  $x$  that if it were to have  $P$  then it would have  $Q$ . The price we pay for retaining these ideas is a commitment to *de re* necessity.

## XI

We have arrived at the following principles:

- I. If  $P=Q$  then  $(\forall x)(x \text{ has } P \equiv x \text{ has } Q)$
- V. If  $P=Q$  then  $(\forall x)N(x \text{ has } P \equiv x \text{ has } Q)$
- VII. If  $P=Q$  then  $(\forall x)(\forall e)[x$ 's having  $P$  is causally relevant for  $e$  (or conversely)  $\equiv x$ 's having  $Q$  is causally relevant for  $e$  (or conversely)]
- VIII. If  $P=Q$  then  $(\forall x)(\forall e)N(\text{the biconditional in the consequent of VII})$ .

For properties that items have at or during a time these principles should be augmented by (or under-

stood as) principles containing a reference to time (thus II, and corresponding ones for V, VII, and VIII).

What logical relationships exist among these principles? V entails I and VIII entails VII, but not conversely. We have also seen that V does not entail VII since the consequent of V does not entail the consequent of VII. The examples at the beginning of Sect. VIII satisfy the consequent of V but not that of VII. These examples also show that V does not entail VIII since if it did then because VIII entails VII, V would entail VII. The remaining question is whether VII or VIII entails I or V.

Suppose some  $x$  has  $P$ . This fact together with the consequent of VII does not entail that  $x$  has  $Q$ . This means that the consequent of VII does not entail the consequent of I, and hence that VII does not entail I. It is only if we combine " $x$  has  $P$ " and the consequent of VII with some further assumption that we can infer " $x$  has  $Q$ ." Two such assumptions either of which will work are deterministic principles for properties:

- D<sub>1</sub>.  $(\forall x)(\forall P)[x \text{ has } P \supset (\exists e)(e \text{ is causally relevant for } x\text{'s having } P)]$
- D<sub>2</sub>.  $(\forall x)(\forall P)[x \text{ has } P \supset (\exists e)(x\text{'s having } P \text{ is causally relevant for } e)]$

For example, suppose  $x$  has  $P$  and we assume D<sub>1</sub>. Then there is some  $e$  (say  $y$ 's having the property  $R$ ) that is causally relevant for this. But then by the consequent of VII,  $y$ 's having  $R$  is also causally relevant for  $x$ 's having  $Q$ . If so then  $x$  has  $Q$ . (One state of affairs is causally relevant for another only if the latter did or will obtain. See Sect. X.) Accordingly, if we can assume D<sub>1</sub> then given the consequent of VII we can infer the consequent of I. This means that VII together with this form of determinism entails I. But VII by itself does not entail I.

The next question is whether VII entails V, which is tantamount to asking whether the consequent of VII entails the consequent of V. To provide a rigorous answer a fuller account of the necessity operator in V should be given. However, the following informal argument may suffice. Consider some  $x$  which lacks  $P$ , and make the counterfactual assumption that  $x$  has  $P$ . If the consequent of VII entails the consequent of V, then from the consequent of VII plus the counterfactual assumption that  $x$  has  $P$  we should be able to draw the counterfactual conclusion that  $x$  has  $Q$ . But this conclusion manifestly does not follow from the consequent of VII together with the counterfactual

assumption that  $x$  has  $P$ . The same is true if we substitute the consequent of VIII for the consequent of VII. However, in the latter case if we combine the counterfactual assumption that  $x$  has  $P$  and the consequent of VIII with the following principle of determinism (which is stronger than  $D_1$ ) then we can draw the counterfactual conclusion that  $x$  has  $Q$ :

$D_3$ .  $(\forall x)(\forall P)N(x \text{ has } P \supset (\exists e)(e \text{ is causally relevant for } x\text{'s having } P))$ .

Since neither VII nor VIII entails V, or conversely, but V entails I, and VIII entails VII, we are left with Principles V and VIII (and their temporal counterparts) as providing logically independent conditions for property identity. Any argument that these principles provide necessary conditions should be based on an appeal to whatever intuitions we have about correct use of the expression "same property" and on an appeal to examples we regard as paradigms and those we regard as inadmissible. I believe that saying that if two properties are identical then anything is such that necessarily if it has one it has the other and anything is such that necessarily if its having the one causes or is caused by something then its having the other causes or is caused by the same thing conforms with, or at least is not violated by, intuitions we have about "same property." And with these principles we are able to exclude various examples on the dubious list of Sect. I that we want to exclude without having to abandon our paradigm examples.

My conjecture is that these principles also constitute sufficient conditions for property identity. This, however, is much more difficult to establish. It would need to be argued that there are no other logically independent principles required by the use of "same property" and that there are no examples that satisfy these principles but are unacceptable property identity statements. Earlier it was argued that principles of explanation and confirmation, although they might be claimed to conform to intuitions about "same property," yield examples we regard as unacceptable. And I have not been able to find examples that satisfy V and VIII that would generally be regarded as inadmissible property identities. But clearly this is not enough to establish the sufficiency of these principles.

A final comment is in order about what in Prin-

ciples VII and VIII I have called the state of affairs of  $x$ 's having  $P$ . What consequences would follow from treating a state of affairs as an entity over which quantification is possible, e.g., as  $(\lambda z)H(P, x, z)$ , to be read as "the  $z$  which is the having of  $P$  by  $x$ "? Principle VII would then assert the following:

VII. If  $P=Q$  then  $(\forall x)(\forall e)[(\lambda z)H(P, x, z)$  is causally relevant for  $e$  (or conversely)  $\equiv (\lambda z)H(Q, x, z)$  is causally relevant for  $e$  (or conversely)]

But it is a theorem of first-order logic that if  $P=Q$  then  $(\lambda z)H(P, x, z) = (\lambda z)H(Q, x, z)$ , and also that  $(z)(e)(z \text{ is causally relevant for } e \equiv z \text{ is causally relevant for } e)$ . Accordingly, if states of affairs are construed as above then Principle VII is a truth guaranteed by first-order logic alone. It is a special case of the following truth of second-order logic:

If  $P=Q$  then  $(\forall \phi)(\forall x)[(\lambda z)H(P, x, z)$  is  $\phi \equiv (\lambda z)H(Q, x, z)$  is  $\phi$ ].

On the above construal of states of affairs Principle VIII would assert the following:

VIII. If  $P=Q$  then  $(\forall x)(\forall e)N[(\lambda z)H(P, x, z)$  is causally relevant for  $e$  (or conversely)  $\equiv (\lambda z)H(Q, x, z)$  is causally relevant for  $e$  (or conversely)].

But this is not a truth guaranteed by logic, since it is not a truth of logic that if  $P=Q$  then  $N[(\lambda z)H(P, x, z) = (\lambda z)H(Q, x, z)]$ .

It should also be noted that if states of affairs are construed as above then a sufficient condition for their identity is automatically provided. Since " $P=Q \ \& \ x=y \supset (\lambda z)H(P, x, z) = (\lambda z)H(Q, y, z)$ " is a truth of logic, it follows that two states of affairs are identical if the properties involved and the things having them are identical.

There is a dispute between Kim and Davidson concerning the individuation of (what they both call) events. According to Davidson the same event makes "Socrates died in prison" and "Socrates died from hemlock poisoning" both true.<sup>27</sup> According to Kim these sentences describe different events since the property of having died in prison and the property of having died from hemlock poisoning are not the same. If Kim means by an event what I have been calling a state of affairs, there is no real dispute here since the ontological entities being discussed are different. There is the *event* of

<sup>27</sup> See, e.g., his "The Individuation of Events," *Essays in Honor of Carl G. Hempel*, *op. cit.*, pp. 222-223.

Socrates' death which, as Davidson urges, can be variously described as Socrates' death in prison or as Socrates' death from hemlock poisoning. But there is also the *state of affairs* of Socrates' having the property of having died in prison, and this is different from the state of affairs of his having the property of having died from hemlock poisoning (just as the fact that he died in prison and the fact

that he died from hemlock poisoning are different *facts* about Socrates). Socrates' death (the event) is something that occurred rather quickly and painlessly and was witnessed by Phaedo, his having the property of having died (the state of affairs) is none of these things. Accordingly, the above criterion of identity for states of affairs is not incompatible with Davidson's position on events.<sup>28</sup>

*The Johns Hopkins University*

*Received August 31, 1973*

<sup>28</sup> I am indebted to Stephen Barker, Robert Cummins, Dale Gottlieb, Jaegwon Kim, and George Wilson for very helpful criticisms and suggestions. Work on this essay was supported by the National Science Foundation.

### III. HEGEL'S EPISTEMOLOGY

ROBERT C. SOLOMON

THE philosophy of Hegel is still sufficiently disreputable in this country, despite the rebirth of interest in Kant, Marx, and Kierkegaard, that a paper on Hegel typically turns into a defense of his philosophy, and then of its intelligibility rather than its truth. Accordingly, I do not want to begin by claiming that Hegel is doing something ingenious and revolutionary—that is still commensurate with being disreputable—but by showing how he was doing something respectable and defensible. I here want to examine one aspect of Hegel's thought—his epistemology [or perhaps we should say, his "ontology of knowledge," for reasons which will soon become clear]. My aim is to establish this epistemology within the tradition usually traced from Descartes to Kant, where it is eclipsed by Hegel's shadow for a century before being brought back to the light by Moore and Russell. But Hegel took Kant's philosophy to be "the basis and point of departure of modern German philosophy."<sup>1</sup> Accordingly I want to argue that Hegel's epistemology is an attempt to rework and make consistent the key arguments in Kant's "Transcendental Analytic" of the *Critique of Pure Reason*. The focus of my attention will be the Introduction, the first three chapters, the first few pages of "Self-Consciousness" and Section AA(V), "Certainty and the Truth of Reason," in Hegel's *Phänomenologie des Geistes*.<sup>2</sup> But, in addition to the more substantial agreements that I shall illustrate between the two works, there is the further similarity in style—both hastily written transcriptions of ideas and arguments not yet fully digested and formulated, and so what I present here is a defensible reconstruction of their thinking rather than a line by line commentary on the texts.<sup>3</sup>

The basis of the interpretation I wish to defend is this: Kant's great anti-sceptical move in the

"Transcendental Analytic" is the mutual rejection of the Cartesian claim to the epistemological priority of the mental and the empiricist model of knowledge as passive-receptive "representation." In place of both, Kant supplies the revolutionary notion of a *a priori synthesis*, the idea that objects are not simply *given* in experience but rather constituted or synthesized as a necessary condition for experience by the pure concepts of the understanding. Similarly, Kant argues that the idea of necessary connection between events, e.g., causal connection, is not given in experience (thus agreeing with Hume) but is supplied by the Understanding as a necessary condition of every experience. But Kant, like most great revolutionaries, is not quite bold enough to step surefootedly onto the new intellectual ground he had liberated. Even as he attacks the Cartesian epistemological priority of the mental, he still finds problematic the idea that we could know things-in-themselves. And even as he attacks the "myth of the given" of empiricism and the "blindness" of non-conceptualized experience, he hangs onto the empiricist notion of "impressions" (*Empfindung*) which are given in atomistic bits and then synthesized to give us objects. And as he argues for the active role of the Understanding and the Imagination in perception, Kant retains the conservative belief that there is but one set of categories and consequently but one possible conception of the world. The move from the idea that *we* supply the categories by which objects are synthesized to the idea that we might supply *other* categories doesn't entice him. But the idea that necessary truths are not in turn necessary (a denial of Becker's reiteration of modalities theorem in the material mode) did not bypass Fichte, who saw there the germinal idea of Kant's philosophy. Unfortunately, Fichte was not equal to the task of

<sup>1</sup> *The Science of Logic* tr. by Johnston and Struthers (London, 1929), I: 44, and "modern German philosophy" in Hegel means "Hegel."

<sup>2</sup> All quotations from Hegel's *Phänomenologie des Geistes* are from Johann Schulze, Ed. (Berlin, 1832). The translations are my own, with reference to Baillie's translation (*The Phenomenology of Mind* [New York, 1967]). All references to the *Phänomenologie* are given in the text, first to the Schulze edition, then to Baillie's rendition (e.g. "(59 B131)"). All references to Kant's *Critique of Pure Reason* are to the Kemp-Smith translation (New York, 1966), second edition, in parentheses in the text prefaced by 'B'.

<sup>3</sup> I have provided such a commentary, upon which my argument here will ultimately depend, in my forthcoming "In the Spirit of Hegel."

redoing and "systematizing" Kant's philosophy, and so the challenge passed to Hegel. Working a quarter century after the publication of the revolutionary *Critique*, whose ideas were now as cool and established as the banal slogans of the great political revolutions of the same period, Hegel, without Kant's timidity and with newfound transcendental Arroganz, pursued to the limits Kant's ideas of a priori synthesis and his rejection of the idea that objects are simply *given* in experience.<sup>4</sup>

The interpretation falls into three parts: first, I provide a cursory and barely adequate summary of the three chapters on "Consciousness" in Part A of the *Phenomenology*. Secondly, I offer a semiformal account of the arguments which I find contained, presupposed, and suggested in these chapters together with the preceding Introduction and succeeding transition to later sections of the *Phenomenology* on "Self-Consciousness" and "Reason." Thirdly, I attempt to support this account by establishing Hegel in a methodological tradition within which I think his epistemological claims can best be understood.

## I

Each of the first three chapters of the *Phenomenology* presents us with what Hegel calls a "form of consciousness;" together, these three are entitled "Consciousness." It is important to stress from the beginning that "consciousness" here does not refer to the specifically "mental," but is rather more like "knowing something." "Knowing something," after Hegel has rejected the idea of "knowledge by acquaintance" in the first chapter means "*knowing that something is the case.*" Each "form of consciousness" is a different philosophical analysis of knowing. In these chapters, Hegel does not simply "take up" a form of consciousness, as is often suggested, but examines and criticizes each of them as *theories*. In the text, we often find reference to "we who are watching the process" or "we who perceive the contradiction." In other words, Hegel's text often says something of a form of consciousness which it could not say of itself.

"Consciousness" is an examination of theories of knowledge which themselves are reflections on knowing, but not self-reflective. "Sense-Certainty," "Perception," and "Understanding" are each a philosophical theory of knowledge. The knowledge they theorize about—even if it should be "theory-laden"—is not self-reflective, and neither are these theories themselves self-reflective. It is only with Hegel that philosophy is forced to be fully self-reflective.<sup>5</sup>

The common link of all forms of "Consciousness" is the fact that the object is taken as "more essential" than our knowledge in the sense that "the object exists, indifferent to whether it is known or not, . . . but the knowledge is not if the object is not" (75/B151). Accordingly, the most primitive of these "forms of consciousness" is "Sense-Certainty," which Hegel describes as,

That knowledge, which is in the first place or immediately our object of concern, can be no other than that which is itself immediate knowledge, knowledge of the immediate or what is. (73/B149.)

It is "mere apprehension (*Auffassen*) free from conceptual comprehension (*Begreifen*)" (149). In "Sense-Certainty," it is suggested, as in naive realism but also as in many empiricisms, romanticisms, intuitionisms, mysticisms, etc., that knowledge of objects begins with *acquaintance*. An object *has* properties, but it is first of all a mere "This," that of which the properties are predicated. Included here is what recent epistemologists have called "the myth of the given," Kant's "blind" intuitions without concepts, and what recent metaphysicians have called a "bare particular." But, Hegel argues, there can be no knowledge that is not knowledge by description as well. "Sense-Certainty" is the barest form of "the myth of the given," and Hegel's argument against it is intended as a general argument against all philosophical claims—whether of naive realism or sophisticated phenomenalism—which argue that it is possible to identify the objects (or sensations) of one's immediate acquaintance without presupposing categories and descriptions. In a familiar Wittgen-

<sup>4</sup> The Copernican revolutionary had died two years before, but one can imagine his ire had he had the opportunity to read Hegel. We recall his "Open Letter" on Fichte (Aug. 7, 1799): ". . . I am so opposed to metaphysics, as defined according to Fichtean principles, that I have advised him in a letter to turn his fine literary gifts to the problem of applying the *Critique of Pure Reason* rather than squander them in cultivating fruitless sophistries. . . ."

<sup>5</sup> In contemporary terms, Hegel does not presuppose the truth of any "knowing entails knowing that one knows" thesis. Nor would he hold even the weaker thesis that "knowing entails believing that one knows." All such theses, Hegel would rightly argue, presuppose a confusion between the Kantian thesis that reflection on knowledge shows that, if we know something (prereflectively), that we can know (reflectively) that we know, and the very different thesis that knowledge itself is necessarily reflective.

steinian argument, Hegel shows that a "this" presupposes an understanding of "what," that identifying a particular presupposes being able to describe it in universal terms. Since the idea of "the given" in philosophical theories depends upon the claim that what is given can be identified independently of and prior to any particular description of it, Hegel, in denying this, undermines its role in epistemology. It must be stressed that Hegel does not deny that knowledge has its sensuous component. He insists, as much as Kant, that it is sensibility that allows us to distinguish between the objects of experience and objects which are only "thought" and merely possible. But, Hegel argues, there is nothing to be said about sensibility insofar as it is "given" and precedes conceptualization. Accordingly, it drops out of the dialectic from the very beginning. To understand knowing, Hegel continues, is to understand how we *describe* the objects of our knowledge. Knowing surely depends upon the sensible but not upon "given" non-conceptual sensible particulars. Our analysis of knowing must turn to universals, the subject matter of the remainder of Hegel's philosophy.

The second chapter is called "Perception," which

takes what is given to it as universal. As universality is its overall principle, so too are its moments distinguished in their immediacy [that is, as universals], I is a universal, and the object is a universal. (84/B162.)

"Perception" is an attempt to replace the ineffability of the "bare particular" of "Sense-Certainty" with a conception of knowledge which suggests that *all* that we are acquainted with are properties ("sensuous universals") and that an object is nothing but a unity of properties. Within this quasi-Leibnizian view (but some empiricist sense-datum theories would also fit here) Hegel also finds it necessary to defend the principle of the identity of indiscernibles, the only way to avoid the fatal conclusion that there might be two objects with all their properties in common and which therefore could not be individuated on this account of knowledge. But, even with this principle, "Perception" will not do. How is it, Hegel asks, that

we take properties to be properties of anything particular? To answer this problem, Hegel introduces the form of consciousness called "Understanding," which "takes the unconditioned universal as the *object* of consciousness" (100/B180). The unconditioned universal (the category of substance) is to be contrasted with "conditioned universals" or properties. The Understanding introduces the notion of "Force," (Leibniz' dynamic version of "substance") as the "supersensible medium" which *has* properties (its "manifestations") and which explains our ability to individuate objects and re-identify them over time and through change. With the conception of supersensible forces and substances, Hegel traces the classical retreat from the "given in experience" of both Empiricists (Locke) and the Rationalists. But the same problem of individuation repeats itself; how do we individuate substances and forces? And, if we are *given* appearances but intellectually *know* this "supersensible world," how are these two related? (Kant's rationalistic problem in the *Dissertation* of 1770 and the *poltergeist* of the *Critique*.) With the former question, Hegel is still in league with Kant, but with the latter question, he begins to turn against him. Not only the idea of "substance," but any idea of the supersensible raises the same problems. In a perplexing but delightful counterexample of sorts, Hegel creates an "inverted" (*Verkehrte*) supersensible world, in which it is suggested that every proposition true of the apparent world corresponds to its *opposite* in the "real" world (black is white, good is evil). And so Hegel rejects, by making fun of, Kant's "things-in-themselves" as distinct from phenomena.

Leaving "Understanding," we also leave "Consciousness" for the famous section on "Self-Consciousness." This is now called "the native land of truth" and "a new form and type of knowledge" (131/B219). With this move, Hegel rejoins Kant and insists that any adequate theory of knowledge takes the contribution of the knower as fundamental. In other words, *reflecting* on (our knowledge of) objects philosophically ultimately must become self-knowledge.<sup>6</sup> At this point (for reasons we cannot discuss here) the dialectic takes

<sup>6</sup> This claim must be distinguished from a superficially similar trivial claim and two false claims. The trivial claim is that knowledge is always predicable of somebody (there cannot simply be knowledge, there must be somebody's knowing). The first false claim is that to know something is to know something about yourself (e.g., that you know something or believe something). But, as I argued above, Hegel need not and does not hold that knowledge itself is reflective: his claim is that *reflection* on knowing shows that the conditions for knowledge are provided by the knower. (See above.) The second false claim is that one can ultimately only know himself. But the very point of Kant and Hegel's joint polemic is to prove the contrary, that what we know are objects "outside" of ourselves. (This and the earlier similar point stimulated by Paul Ziff.)

a turn into what appears to be primitive social philosophy and the concept of "freedom." Hegel's theory of knowledge as such re-emerges in the next section with "Reason," where "the individual consciousness is inherent absolute reality" (174/B272) and "is the conscious certainty of being all reality" (175/B273). Here Hegel leaves Kant for Fichte. His concern is "Idealism" and Hegel is concerned to protect the gains of "Transcendental Idealism" (Kant and Fichte) from the confusion (which virtually destroyed Fichte) between Transcendental Idealism (in Fichte, "Ethical Idealism") and Subjective (or what Hegel calls "Empty" or "Unintelligible") Idealism (175f/B273).

No doubt the language of this argument is unfamiliar today and its structure perplexing; in fact, it is not obviously an *argument* at all. But everyone who knows nothing else of Hegel knows that the above progression is a form of "dialectic" and that the stages are arranged in ascending order approaching what Hegel pretentiously calls "Absolute Truth." But how is this an argument? Where is the "epistemology?" Why the opaque references to Locke, Leibniz, Kant, and Fichte? Why these particular forms? And why this particular progression?

Notice that Hegel is cutting the philosophical pie in an unusual way. Unlike Kant, he does not take the competing epistemological claims of the empiricists and the rationalists as the antitheses for him to *aufheben*.<sup>7</sup> Rather he sees a series of competing ontological claims which are, as often as not, shared by the two epistemological traditions. (Locke, Leibniz, and Russell, for example, would have to be included at every stage.) This is why we began with the parenthetical remark that epistemology for Hegel is "the ontology of knowledge." One must distinguish two very different epistemological problems, one of which Hegel took Kant to have already solved beyond question. This is the empiricist "constitution" problem—how do our sensations or "impressions" become organized into objects? Hegel accepted Kant's answer without qualification: *we* so organize our experience into objects, and so it does not even make sense, Kant proved, to talk about experience except in terms of our experience of objects. Accordingly, Hegel does not talk about it in any other terms, and the familiar empiricist epistemological problems con-

cerning the relationship between experience and objects are not even mentioned by him. However, there is a second problem—also part of Kant's "reply to Hume"—which arises independently of the "Constitution problem." Assuming that we are conscious of objects, how can we know that the objects of which we are conscious are "objective"? How can we know that they exist independently of our awareness of them, that the space they occupy is "real" space, that they are not just productions of *my* consciousness? To this problem, Kant gave a promising but incomplete answer. The categories by which I constitute my experience are such that to be conscious of objects is *ipso facto* to be conscious of objects as independent, as "real" in "real space," and not just *mine*. This answer, too, Hegel accepts without question, so far as Kant had presented it. But there is still a question, left unanswered by Kant, concerning the *ontological status* of the objects of knowledge. Kant had reduced all ontological questions to epistemological problems and so used his answer to the "constitution problem" as an answer to the ontological questions as well. The mutant and subtle scepticism of Kant's notion of "things-in-themselves" is but a symptom of this ontological neglect. And so Hegel, accepting Kant's epistemological arguments so completely that he does not even consider it necessary to repeat them, goes on to provide the ontological counterparts of those arguments. What is it to know something? We need not even consider the narrow problems raised by the empiricists in their peculiar reply to this question. Hegel does not feel it necessary to give another "reply to Hume." He does not feel it necessary to restate Kant's theory of "constitution" or "a priori synthesis." But he does find it necessary to return to the epistemologies that Kant rejected to further purge them of shared ontological claims which Hegel wants to reject by similar "transcendental arguments."

The first three "forms of consciousness" are each one of these shared ontological claims: first, there is the naive idea of a "bare particular," knowable by "pure acquaintance"—whether such objects of acquaintance are formless sense-data or impressions, a bare particular or substance, "Being-in-general" or "Absolute Ego" as known by "intellectual intuition" (Jacobi, Fichte, and Schelling; also cf. Pt. 1 of the *Logic*). Second, there is the idea that objects are merely "clusters" of

<sup>7</sup> The best definition I know of "*aufheben*" has been proffered by the hardly Heglian philosopher Frank Ramsey in his *Foundations of Mathematics* (pp. 115–116); "the truth lies not in one of the two disputed views but in some third possibility which has not yet been thought of, which we can discover by rejecting something assumed as obvious by both the discussants."

properties (or sense-data, which are here conceptualized and describable, not "This" but "red and square"), again a thesis which can be found both in the empiricists (e.g., Berkeley) and rationalists (e.g., Leibniz). Third, there is the idea of supersensible substance or force, a thing-in-itself that in some sense lies "behind" properties and particulars and accounts for them. Substance is recruited to fill in for the rejected bare particulars, but without the naive claim of knowledge by acquaintance: substances can only be known by "the Intellect" (for the rationalists), or inferred (e.g., by Locke). Against those commentators who loudly complain about the arbitrariness of Hegel's "Dialectic" (e.g., Findlay, pp. 55, 69-73, 93, 107, 113, 118, 143, 147), I would want to argue that the logical progression here, which focuses on the single but central ontological problem of individuation of particulars, is as rigorous as one would expect even from contemporary philosophers. There is the natural but naive idea that "objects are simply *there*, and we know things because they are there," against which Hegel argues that one cannot know particulars (by acquaintance) without knowing universals (by description as well as acquaintance). Accordingly, it is suggested that what we are acquainted with is not particulars but universals, that is, properties. But Hegel argues, one cannot individuate particulars without already knowing *of what* a "cluster" of properties is predicated. And so, the idea of knowledge by acquaintance yields to the suggestion that *we* predicate properties (which are still known through acquaintance) of particulars by reference to a supersensible substance which *has* those properties. But then, stuck again, for how does one individuate supersensible substances without already presupposing individuation of (sensible) objects. Accordingly, the next move is Kant's move (from his *Dissertation of 1770* to the *Critique*); he changes the philosophical notion of "Understanding" (or "Intellect") such that its pure concepts apply, not to the supersensible things themselves but to appearances. Objects are not *given* by acquaintance with particulars or clusters of properties, but neither are properties simply given as properties of something (either a bare particular or a supersensible substance). *We* (in some sense) are responsible for *producing* the objects which we are apparently given. And with this, Hegel's dialectic moves from mere consciousness of objects to self-consciousness to the notion of *Reason*, that our activities are as essential to the object of knowledge as the object is to our knowledge. (Cf.

consciousness above.) Hegel has arrived at Kant's conclusions through similar reasoning, but along the route of an epistemo-ontological argument in the place of Kant's epistemological "Deductions." We might here add, also against many commentators (again e.g., Findlay, pp. 67-68, 78-79, 93) that it is clear on the above interpretation that there are a number of possible routes through the dialectic to any given point, including (especially) its end-point ("Absolute Truth").

## II

The section on "Consciousness" is preceded by a brief Introduction to the entire work which can function for us here as a guide to just these chapters. The focus is on Kant and the empiricists, although, as usual with Hegel, no one is mentioned by name. Kant had complained that philosophers were trying to settle metaphysical disputes concerning "things-in-themselves" before they examined the possibility of gaining any such knowledge. Accordingly, with the empiricists, he reduced all ontology to epistemology—what there is depends upon what we can—and must—know. Consequently, there can be no knowledge of "things-in-themselves" (no "metaphysics"). Hegel agrees that ontology must be limited by epistemology, but he adds that Kant is mistaken if he believes that epistemology can avoid metaphysical as well as ontological commitments. [We see precisely the same dispute repeated a century later in Heidegger's rebellion against Husserl.] In this Introduction, Hegel argues this point by accusing Kant (and the empiricists) of employing an unexamined metaphor, of knowledge as an "instrument," as a "medium" through which we come to know objects:

It is a natural idea that, before philosophy would get to the subject matter itself—namely to the actual knowledge of what is in truth—it is necessary first to understand that knowledge itself, which is considered to be the instrument through which one takes possession of the absolute itself, or as the means through which one discovers it. (59/B131.)

But instruments "mould and alter" objects and media refract and distort the "light of truth" that passes through. (Cf. the "spectacles" and "meat grinder" metaphors that have been suggested by two of Kant's leading British commentators, Kemp-Smith and Weldon.) Consequently, Kant finds himself with two different senses of "object" and "truth"—as known by us (as "phenomena") and as they are "in-themselves" (as "noumena").



Hegel does not attack Kant's phenomenon-noumenon distinction; he undermines it, finally reducing it to an absurd joke at the end of the chapter on "Understanding." And again, he does so by using Kant's best arguments against him. Just as Kant had argued that traditional scepticism and idealism resulted from an indefensible distinction between experience and the objects of knowledge, Hegel now argues that Kant committed a similar error in his distinction between the objects of knowledge and the objects-themselves. Against traditional scepticism, Kant had leveled his "Refutation of Idealism." But Hegel does not see scepticism as a problem; rather he sees seeing scepticism as a problem as a problem;

We should beware that this fear of erring is already the error itself. Actually, this presupposes something and indeed many things as truth; it presupposes the same images of knowledge as an instrument and a medium, and a distinction between ourselves and this knowledge. (61/B132-133.)

Kant had paid too little attention, according to Hegel, to the role of philosophical *critique* in the determination—as well as the analysis—of the Understanding. It is this neglect that leads him to his "phenomenon-noumenon" thesis and his "new" scepticism. Against all forms of "idealism," Kant had argued that the Understanding not only understands but contributes to experience. Objects are not inferred from but are the very structure of experience. And now, against Kant, Hegel argues that it is similarly a mistake to suppose that philosophical criticism only criticizes rather than contributes to our understanding. Philosophy is not only reflection upon but the structuring of our knowledge. Kant's mistake, according to Hegel, is his idea that the understanding and empirical knowledge are independent of philosophical criticism, whose only function is to justify their concepts and not to determine them. Accordingly, the *Critique* succeeds in this justification only by allowing philosophy (which is, for both Kant and Hegel, a matter of *reason* rather than *understanding*) to entertain the impossible notion of things as they are apart from our understanding.

Kant's "reply to Hume" was that knowing is a self-confirming activity. Hegel's reply to Kant is that philosophizing about knowing is also. As he states this,

Consciousness is for itself its own concept . . . Consciousness furnishes its own criterion of itself, and the inquiry [of philosophy] will thereby be a comparison of itself with its own self.

This central Hegelian claim, that knowing and philosophizing about knowing are self-confirming activities, does not entail the absurdity (with which both Hegel and Fichte were often charged) that all our knowledge is our own creation. There is nothing here that denies the possibility of error or ignorance regarding empirical propositions (let us call any such proposition '*P*'). But any proposition *P* is *presupposed* by a number of other propositions ('*C*'), which are not only required in order for *P* to be true but in order for *P* (or not-*P*) to be intelligible.<sup>8</sup> In any dispute over *P*, *C* is taken as unquestionably true. But there is, according to Kant, a class of propositions such that they are presupposed (in the above sense) in every empirical dispute. Let us call such propositions '*C\**'. One such proposition would be what Kant calls the *causal maxim*—the a priori principle that every event succeeds another in accordance with a rule. The claim that knowledge is self-confirming maintains that it is impossible to be mistaken concerning propositions *C\**, the a priori maxims which are presuppositions of *all* particular empirical beliefs. But the reason one cannot be mistaken in believing *C\** is not to be attributed to any peculiar indubitability of *C\** or the inconceivability of not-*C\**. Rather, belief in *C\** cannot be wrong because it is *taken* as unquestionably true in every empirical dispute. In other words, one *makes it true* by believing it and not questioning it.<sup>9</sup> It serves functionally as a rule; a rule of inference in argument, a rule of understanding in experience. A rule, however, can always be stated as an axiom—in which case it will appear as a necessary truth. For example, Kant's causal maxim might appear either as a rule, "Take every event to be . . .," or as an axiomatic principle of experience, "every event

<sup>8</sup> It must be insisted that the same propositions might be either *P* or *C* depending upon the theoretical role they played in a particular context. *C* is a proposition which must be taken to be true if *P* is to be considered either true or false. In the terms of the current controversy in linguistics, the conception of "presupposition" I am employing here is "pragmatic" rather than "semantic." I do not believe that my argument here turns on these issues. Cf. Stalnaker, "Pragmatic Presupposition," unpubl.

<sup>9</sup> The obvious analog here is J. L. Austin's notion of a "performative," but the more apt comparison is Carnap's discussion in Appendix A of *Meaning and Necessity*.

has a cause." Such principles do not have the status of necessary truths because of any conceptual necessity (although principles such as  $C^*$  are, according to both Kant and Hegel, *concepts*. But, for them, concepts are *rules*, even if Kant, like Hume but unlike Hegel, sometimes slips into a more psychological characterization.)  $C^*$ , as a rule, is the product of a decision, not a fact about the world or a description of the subject's activity. It is a basic presupposition, a self-confirming stipulation that provides a framework for empirical knowledge.

It is in this sense that both Kant and Hegel insist that knowledge is self-confirming. The key notions of "a priori synthesis" in Kant and the "universal in action" in Hegel converge in their mutual insistence that certain basic concepts of the understanding of experience are contributed by the subject. Their mutual antagonist is Locke and the empiricists who took knowledge and understanding to be a matter of correspondence of a given representation to the object represented. Against this, Kant and Hegel are arguing a material mode equivalent of a principle of great importance in contemporary philosophy; that there are no concept-free representations. What we can know is determined by the concepts—and that means by the language—we provide to describe what we know. In Quinean terms, "experience underdetermines theory;" "the central fundamental theoretical commitments in the organization of experience are one's own contribution."<sup>10</sup>

If  $C^*$  is a necessary truth because it acts as a rule and therefore remains unquestioned in empirical disputes, it does not follow that it cannot be questioned elsewhere—for example, in philosophical criticism. Kant, of course, recognizes this, and states clearly in mid-*Critique*

... while the understanding, occupied merely with its empirical employment, and not reflecting upon the sources of its own knowledge, may indeed get along quite satisfactorily, there is yet one task to which it is not equal, that, namely of determining the limits of its employment ... (B 297.)

Now Hegel agrees with Kant—he was taught by Kant—that it was a mistake for Locke, *et al*, to assume that the Understanding only understood—rather than determined—the knowledge it was apparently "given" through the senses. But now

Hegel goes on to make a similar point against Kant, who makes the same mistake with regard to philosophical criticism. Philosophy does not merely recognize but determines the structures it apparently "finds" in the activities of knowledge. In other words, Kant had argued that  $C^*$  counts as a necessary truth because it functions as a rule defining empirical considerations which cannot be challenged in the course of such considerations. (In seeking a causal account of  $P$ , one cannot even consider the possibility that there might be no such account.) But in a philosophical discussion,  $C^*$  does not function as such a rule, and consequently does not appear as a necessary truth, and can subsequently be challenged to see whether it is a "truth" at all. Within philosophy, it can be shown that  $C^*$  has a presupposition  $Z$ , which it shares with not- $C^*$ .  $Z$  would be a "transcendental judgment," i.e., a judgment *about* another judgment which is taken to be a necessary truth; e.g., B 80. For example, the causal maxim ( $C^*$ ) has as its presupposition the proposition that our experiences are necessarily ordered in time. Kant so argues in the "Transcendental Analytic." But the denial of the causal maxim (not- $C^*$ ) would also take this proposition as a presupposition. Leibniz, for example, rejects the causal maxim but would insist, as a presupposition of that rejection, that our experiences are necessarily ordered in time. Kant, who holds  $C^*$ , necessarily holds  $Z$ . But Leibniz, who holds not- $C^*$ , finds it equally necessary that he hold  $Z$ . Parodying Quine and paraphrasing Hegel, we may say that necessary truths underdetermine our philosophical theories. Empirical considerations are not capable of deciding between  $C^*$  and not- $C^*$ , since these principles in turn are responsible for defining empirical considerations in general. Hegel's argument, then, is that whether or not one accepts a given  $C^*$  is a matter of some dispute, but not an empirical dispute, and that the philosophical model one accepts in part determines the rules for empirical knowledge one can accept and thereby determines the objects one knows. If our experiences depend upon the way we understand them, our understanding in turn depends upon the way we reason.

Kant was instrumental in rejecting the empiricist notion of the *given* in experience. But he replaced it, according to Hegel, by introducing a *given* in the

<sup>10</sup> "Two Dogmas of Empiricism," in *From a Logical Point of View*, pp. 42-43. Of course, the theoretical commitments here would include, for Hegel, philosophical commitments as well as the ontological commitments of scientific (empirical) theories. We might also add that Quine, still an empiricist at heart, takes the "observational edge" to be the key to conceptual change. For Hegel, always a rationalist, it is rather the "form of consciousness" that takes priority, not the evidence.

Understanding.<sup>11</sup> Many commentators have balked at the source of Kant's collection of categories, the over-ripe fruits of the pretentious and antiquated logic of the day. But Hegel argues further, not against the source or even the particular list of categories Kant provides, but against the very kind of a fixed and determinate list of a priori concepts without regard to the kinds of reasoning that is going on about them. Differing conceptions of the understanding will in turn produce different conceptions of the concepts of the Understanding; in other words, different philosophers will produce either different categories or different views of the categories, and a philosophical critic who reflects on these differences will be forced to give up the concept of "understanding" as it had been employed by Kant and others as a faculty with a determinate set of concepts. For example, Locke, Spinoza, and Leibniz, or Plato, Aristotle, and the early Chinese philosophers are not just giving different *accounts* of similar experiences and the same activities, some of which are more nearly "correct" than others. When Hegel says, in the Preface to the *Phenomenology* and elsewhere, that we should see the "difference of philosophical systems as the progressive development of the truth" (4/B68), he means more than the obvious and patronizing consolation that each philosophy has *something* right about it. In addition, he means, in the sense that we have outlined above, that each philosophy is self-confirming, and that their apparent contradiction can only lead us ("perceiving the contradiction") to conclude that it is the self-confirmation itself, not the *contents* of the various systems (i.e., their categories), that is the *truth* Hegel is talking about. The ("absolute") truth is—there are alternative and equally valid if not equally adequate "one-sided" ways of looking at the world of human experience: the basic concepts of our understanding are determined in these various theoretical stances. In the *Phenomenology* these are called "forms of consciousness."

Accordingly, Hegel rejects Kant's concept of the Understanding and concentrates his attention on Reason. The difference between the two is not, as in Kant, that the former applies concepts to experience while the latter does not; for Hegel, both Understanding and Reason apply their concepts *only* to experience, but the former in an unreflective, rigid, and mechanical way, the latter in a self-reflecting, free-wheeling manner that takes into

consideration every possible coherent alternative system of conceptualization of experience. Understanding is essential for shopkeepers, but, in philosophy, Understanding ("which, itself finite, can understand only the finite" (*Encyclopedia* 28)) is a reactionary idea, the equivalent of bourgeois morality in a revolutionary society, or Newtonian mechanics in a problem in quantum physics. It is a piece of efficient yet gross machinery, adequate to gross tasks, but not equal to the critical tasks of philosophy. As philosophies change, reason "develops," and the basic concepts of experience change also. And experience changes with the concepts of experience. It is Hegel, before Marx, who insists that the point of philosophy is to change the world, although he would add that to understand it *is* to change it.

### III

The Cartesian dictum might better be expressed, "Hey, there goes Edna with a saxophone!"—Woody Allen, *Getting Even*.

It may be that there are presuppositions  $\mathcal{Z}^*$  of all philosophical discussion which occupy a logical role similar to  $C^*$  in empirical discussions. For reasons which will become apparent shortly, I do not believe that there are any such, but Hegel—with most philosophers of his tradition—did believe in one such  $\mathcal{Z}^*$ —the peculiarly necessary "fact" of (my) consciousness. This is famously the starting point of Descartes' Meditations, demonstrably the premise of Kant's transcendental deduction, and arguably the beginning of Hegel's entire system of philosophy. But  $\mathcal{Z}^*$  is not *simply* an ultimate presupposition, nor a premiss of these philosophical arguments. If it were, we could not possibly make sense of the fact that three of Europe's most brilliant thinkers took  $\mathcal{Z}^*$  to be the starting point of their philosophy. If  $\mathcal{Z}^*$  were simply a principle which was presupposed by every proposition,  $\mathcal{Z}$ ,  $C^*$ ,  $C$ , and  $P$  (Kant's "regressive" argument of the *Prolegomena*) then the attempt to argue from  $\mathcal{Z}^*$  to  $C^*$  (Kant's "progressive" argument of the *Critique*) would amount to the elementary logical fallacy of affirming the consequent. There is something most peculiar about  $\mathcal{Z}^*$ .

Kant's brilliant if obscure *Transcendental Analytic* is a turning point *within* an epistemological tradition that includes Hegel, which began with

<sup>11</sup> Cf. Nietzsche (*Beyond Good and Evil*, §11), "Why should we believe such judgments to be necessary?"

Descartes and Locke, came to Kant through Leibniz and Hume, reached Hegel from Kant and Fichte. Basic to this tradition is a philosophical method or discipline which I shall call "methodological solipsism." (I borrow the term from Jonathan Bennett, who also helped me clarify some of the issues raised here.) Methodological solipsism (MS) is what I consider to be the defining characteristic of "Cartesianism" broadly conceived, and it was adopted by not only traditional Cartesians but also most empiricists, Kant, Fichte, Hegel, and, prominently, all those contemporary philosophers who claim to follow "the phenomenological method," which I would argue to be a tedious version of MS. One might first characterize MS by saying that it is a first-person philosophical orientation and consequently restricts the kinds of questions one can ask and the types of appeals one can make. In MS, every philosophical problem must be construed as *my* peculiar problem. The question, "What is knowledge?" must be recast as "What is it for me to know something?" I must determine what is justifiable for *me* to believe according to evidence which *I* have. Similarly, the problem of other minds becomes, "How can I know that other persons exist?" Needless to say, there can be no appeal to other persons, to "common" sense, to theories or models which I cannot justify for myself. MS is the mother of the ego-centric predicament. But these are not identical, for it is the presumption of the philosophers who endorse this program to demonstrate that *our* common knowledge of the physical or natural world can be understood and defended only *in terms of* MS. MS, we can now say, is what lies behind the problematic  $Z^*$  of our argument.

I know this is a markedly unfashionable program in this country these days, since Wittgenstein in particular, but I believe that without it we cannot understand Hegel—or Kant or Descartes, Leibniz or Fichte, or, perhaps much of empiricism, which is a perversion of the MS program. It may be that the entire methodological solipsist program is a dead end, but I would argue that without it the awful pinch of the classical problems of knowledge isn't imaginable. Once we give up MS and "naturalize" epistemology (in Quine's and Nietzsche's

phrase), these problems do not arise in any compelling way, as all the propositions which troubled the MSist are *assumed* to be true in the very statement of the problem. ("How do *we* know that there are other minds?")

I take MS to be the greatest contribution of Cartesianism—its demand for epistemological autonomy and intellectual rigor. But in substance, Cartesianism went off the track virtually at once. Even aside from Descartes' notorious retreats and circles, the method itself took a wrong turn as soon as he stated it. From the first-person stance of MS, Descartes asserted that the mind is better known than the body. Accordingly, MS became the attempt to move from immediately known mental experiences to objective reality. In fact, I shall argue, not only does MS not entail any such "phenomenalism," but it provides overwhelming grounds against any such theory. But MS was turned into a phenomenalist program by most Cartesians, by the British empiricists, and by many philosophers today. Nelson Goodman, to choose but one illustrious example from many, gives the distinction between phenomenalist and physicalist models a prime place in his *Structure of Appearance* (although, while he opts for the former, he rejects the claim of epistemological priority to either). Similarly, the sharpest critics of phenomenism, e.g., W. V. O. Quine (in *Word and Object*) begin by making the same distinction, but then reject one of its terms.

Cartesianism as I want to pursue it, as MS, does not need and ought to abandon phenomenism and the phenomenism-physicalism distinction. Undermining both was one of Kant's greatest contributions to the Cartesian tradition (which was by then so entrenched that it was not even thought of as a tradition). Kant himself remained a Cartesian dualist of sorts, but he was not a phenomenist.<sup>12</sup> It is false, he argued in the Transcendental Analytic (and even more explicitly in the "Refutation of Idealism" in the second edition of the *Critique*), that the mental is better known than the physical. The very existence of a unified self-consciousness presupposes the existence of physical, i.e., spacial objects. What we know are physical objects, not our own sensations.

<sup>12</sup> This claim would appear to contradict the findings of several important commentators, notably J. Bennett in his *Kant's Analytic*. The contradiction is only nominal, I would argue, but will not here, due to small but significant differences in the use of the term "phenomenalism." "Phenomenalism," as I am using that term here, is the thesis that there exist discreet atoms of experience identifiable independently of identifications of physical objects (leaving open questions of how these are identifiable and what the relationship is between experience and physical objects).

But where Kant, betraying his empiricist influences, seems to believe that there were raw sensory materials which, by their very nature, could not be examined as such, Hegel more consistently argues that this very notion is incomprehensible. One of the consequences of the first two chapters of the *Phenomenology* is that any notion we might have of sensations or sense-data (particulars or properties) must be parasitic on our prior knowledge of physical objects.

Phenomenalism is to be distinguished from MS: There is nothing in the first-person method of MS that requires introduction of peculiarly mental entities or events. The objects of scrutiny in MS are physical objects, or perhaps we should say putative physical objects, not experiences. And if MS typically (but not always, cf. Hegel) considers doubts concerning the existence of physical objects, it does not follow that *experience* is better known than objects. Within MS, as in everyday discussion, our talk is physical object-talk. The notion of "experience" can be brought in, if at all, only in those special cases (toothaches, migraines, visual flashes, perhaps hallucinations) where physical object-talk breaks down. Even here, introduction of the notion of "experience" raises difficult problems, particularly within the MS standpoint. (I take this to be the thrust of the "private language arguments" that have been squeezed out of the later Wittgenstein.) But to introduce the notion of "experience" as a general conception of the objects of consciousness is as controversial and counterintuitive within the MS position as in everyday conversation.

In fact, MS excludes phenomenalism. Consider Macbeth at the banquet table watching the procession of Banquo and his relatives. For Macbeth, any conception of "his experience of seeing Banquo" is just an unnecessarily complex way of referring to his "seeing Banquo." It is only the guests who surround him, who do not see what he sees, who could talk significantly about "Macbeth's experience." But they refer to it as "an experience" in such a case just because its objects are not shared by them. For Macbeth, were he a good Cartesian, there could be no appeal to his guests' opinions; so he could have no notion of "his experience" as opposed to simply his "having seen Banquo." In other words, the notion of "experience" comes, not from the first-person standpoint of MS where it is typically identified, but

from the third-person attempt to explain the behavior of others. Quine is right, I believe, but for all the wrong reasons, in his suggestion that experiences and the like are theoretical entities which we supply to explain *other* people's behavior. However, the phenomenologists are correct—but from the *third person standpoint*—in their insistence that such accounts are indispensable. But the Wittgensteinians are also correct that such explanations are not conclusions of analogical arguments from any notions of "experience" drawn from my own case. The notion of "experience" is peculiar to the third-person case; it has no primitive role in the first-person case and therefore has no role in any account of MS.

Properly understood,  $\mathcal{Z}^*$  is a cryptic if not misleading statement of MS, not a "first principle" or premiss, not a peculiarly necessary *fact*. Surely it is not a necessary truth that I exist, or that we exist, or that any conscious creature exists (neglecting certain traditional theological doctrines). The "necessity" of this "fact" of consciousness can *only* appear within the first person case. But even then, it is not as if there is some *principle*, "I am conscious" or "I exist," which becomes a necessary truth; in this peculiar case, the conditions for assertion of a *particular* utterance and the truth conditions for that utterance are identical. This is the point made forcefully but restrictedly by Hintikka, who argues that Descartes' *Cogito* is a "performance." But it must be added that the necessity of  $\mathcal{Z}^*$  lies not only in the peculiar status of some particular utterances.  $\mathcal{Z}^*$  defines the context within which any utterance is permitted.  $\mathcal{Z}^*$  is not our starting point because it is a necessary proposition:  $\mathcal{Z}^*$  appears as a necessary truth because it is our starting point.  $\mathcal{Z}^*$  is nothing other than MS, the insistence that all our philosophical deliberations be carried out in the first person case.

There are well-known and outstanding problems in formulating  $\mathcal{Z}^*$ . Descartes' *Cogito* formulation is crisp, and raises fascinating problems in itself, but it is clearly not correct as a statement of MS. The status of the 'I' is famously suspect, and more than one philosopher of his own time suggested that it did not belong in the formulation at all. For Hegel, notably, the first-person case of MS can only be defended as the first-person *plural*.<sup>13</sup> As the 'I' drops out of the formulation of  $\mathcal{Z}^*$ , the possibility of individuating "consciousnesses" drops out also. But even Hume, in denying the unity of

<sup>13</sup> See my "Hegel's Concept of *Geist*," *The Review of Metaphysics*, vol. 23 (1970), pp. 642–661.

consciousness, is forced to say "I am not aware . . ." Similarly, Kant, who reintroduces the 'I' as a "principle of transcendental unity," wrongly rests his concept of the "I Think" on an unwarranted and unquestioned supposition that "I Think's" can be individuated by individuating persons. But neither can one simply say, as several of Descartes' contemporaries and later Hume and Nietzsche suggested, that  $\mathcal{Z}^*$  is "there is consciousness" (or "there are thoughts"). "Consciousness" here cannot refer, for reasons now familiar, to the phenomenologist notion of "experience" (Hume's conception). "There is consciousness of objects," while compensating for the above faults, carries with it no possible way of identifying "consciousness of objects" as opposed to simply, "objects." Thus James is pushed to the absurd conclusion that "there is no consciousness." To answer this 200-year-old puzzle, Hegel points out that there is no correct formulation of what we are calling ' $\mathcal{Z}^*$ ' as a first principle with the properties of self-certainty etc., demanded by Descartes.  $\mathcal{Z}^*$  is not merely an axiom or premiss but a rule for philosophical discussion. In the *Phenomenology*, it is understood from the outset (as it is also in Kant's *Critique*) that the entire investigation proceeds from the first person case (thus the title, *Phenomenology*, as well as the titling of the first section, "Consciousness"). But  $\mathcal{Z}^*$  does not appear, as in Fichte—whom Hegel attacks on this point, as a "first principle." It is the guiding rule for the entire work, emerging explicitly in the course of the *Phenomenology* as its importance becomes more evident. Thus the notion of "Self-consciousness," the usual formulation of  $\mathcal{Z}^*$  (in Descartes, Kant, and Fichte) does not appear until section B, and Kant's "consciousness in general" taken literally, does not appear explicitly until the final section of the book "Reason." And  $\mathcal{Z}^*$  does not emerge in full glory until the still less-than-adequate formulation as "Absolute Truth." This way of interpreting  $\mathcal{Z}^*$  can, I believe, open up a way of interpreting many of the better known opacities of the *Phenomenology*, notably the confusing discussion of the identity of results and method and subject and substance introduced in the Preface and summarized in "Absolute Truth," but without much explicit help inbetween.  $\mathcal{Z}^*$  is not a first principle but the rules of philosophical inquiry—preventing us at any point from stepping out of MS—the source of the most serious errors of empiricism and, Hegel would argue, of Kant's ultimate mistake—his distinction between noumena and phenomena.  $\mathcal{Z}^*$ , as a rule, might be made into

an axiom at any point, but such a "mere result," without "the process of arriving at it" would be but a "lifeless universal,"

The naked result is the corpse of the system which has left its intention behind. (5/B69.)

$\mathcal{Z}^*$  will appear as a principle as well as maintain its function as a rule only when it is *necessary* for it to do so. In other words, Hegel begins, as his Cartesian predecessors began, by minimizing the effects of their philosophical method and accordingly dealing with purely ontological questions. It is only when he has demonstrated that these ontological puzzles are unsolvable without a little help from  $\mathcal{Z}^*$  that  $\mathcal{Z}^*$  is allowed to make its first assertive appearance in "Self-consciousness." And it is only when the minimal and traditional interpretation of  $\mathcal{Z}^*$  as a *personal* (self-) consciousness fails that  $\mathcal{Z}^*$  becomes increasingly prominent as the "result" as well as the method of *Phenomenology*.

Philosophy does not start from scratch. Of course, everyone knows that Hegel believed his own philosophy to be the extension and incorporation of all the philosophies before him. But in particular, the method he uses, MS, is parasitic not only on two centuries of philosophers before him, but also on everyday language and knowledge. The first person standpoint itself is parasitic on "common sense," since there is no way one could "get into" the first person standpoint of MS without a prior philosophically inadequate but ordinarily adequate conception of himself—himself opposed to others, himself opposed to objects. Here again, Wittgenstein is correct in suggesting that first person ascriptions are parasitic on third person ascriptions. But, to borrow a metaphor from an earlier Wittgenstein, one can always "throw the ladder away."  $\mathcal{Z}^*$ , accordingly, must first be crudely formulated as a rule of method in everyday terms, and this means with some essential reference to oneself as a conscious individual. But this is only a starting point, and further refinements might well end in the elimination of the 'I'—whether implicit or explicit in the initial formulation of  $\mathcal{Z}^*$ . Such an elimination occurs crudely in Hume, and again, far more spectacularly, in Hegel. One might well say that Hegel's *Phenomenology* is basically an attempt to formulate  $\mathcal{Z}^*$  in an adequate way.

Descartes' *Cogito*, Kant's "I Think," and Hegel's *Geist* all function as philosophical rules, ultimate *within*, but *only* within, the first person case. And, of course, a rule can always be employed as an axiom, as a premiss, as Descartes used  $\mathcal{Z}^*$  in his

Meditations, and as Kant has been argued to use the "I think" ("the synthetic unity of apperception") in his Transcendental Deduction (e.g., by R. P. Wolff).

And, taking  $Z^*$  as a premiss or as a first "synthetic" principle, Descartes and Kant (at least according to several commentators, including Kemp-Smith, Paton, and, more recently, Wolff) attempt to argue from  $Z^*$  to  $C^*$ . But the form of this "transcendental argument" is, as we pointed out, seriously misleading. If  $Z^*$  is characterized as the ultimate presupposition of  $C^*$ , then surely  $Z^*$  cannot entail  $C^*$  (assuming that the two propositions are not logically equivalent, which I take to be obvious). The mistake here is thinking that these presupposition and entailment relationships hold (semantically) between two (sets of) propositions, rather than (pragmatically) between a set of utterances  $C^*$  and the philosophical context in which they are placed (as defined by  $Z^*$ ).  $Z^*$  is a presupposition of  $C^*$  only within the MS standpoint; it is MS that dictates the status of  $Z^*$ , not  $C^*$  alone, and so it might be validly argued, *within the context of MS*, that  $Z^*$  entails  $C^*$ . (This is the hidden trick to Wolff's insightful reconstruction of Kant's arguments of the Analytic; MS provides the contextual emphysemes for all his reconstructed arguments.)  $Z^*$  is the "ultimate presupposition" of  $C^*$  only in the sense that it is the ultimate rule for a certain type of philosophical accounting of  $C^*$ . Within the MS standpoint,  $C^*$  (and also other transcendental principles  $Z$ ) appear as "necessary" insofar as they are accepted as rules and not called into question. And Kant, Hegel critically points out, never challenged the principles  $C^*$ ;

The Kantian philosophy allows the categories and methods of ordinary knowledge to remain completely unchallenged. (*Encyclopedia* 60.)

Thus the dispute between Kant, Fichte, and Hegel becomes the following—is there any set of principles with the status of  $C^*$  or  $Z$  that is necessary and indispensable for human experience? According to Kant, there is such a set of categorial principles of the Understanding; according to Hegel, there is no such set, and so for him, strictly speaking, there can be no "transcendental deduction" of a particular set of categories, but only attempts to show what turns on the acceptance of any one set of categories rather than any other. But what is not in dispute between them, or even debatable, is whether some set of  $C^*$  principles or other is

necessary for human experience. This is, of course, Kant's proof in his Transcendental Deduction, which Hegel reaffirms in the first part of the *Phenomenology* and again in the first part of his *Logic*.

The ultimate problem, for Hegel, is  $Z^*$ , which, although it is a rule, is an ultimate rule. One can, we might argue, step out of MS at any time, and see that  $Z^*$  is only a methodological principle, and a dispensible one. But  $Z^*$  was truly the limit of Hegel's investigations, and so  $Z^*$ , for him, remained unquestioned and unquestionable, and so, it became "absolute truth." But it is important to notice what happens if one does step outside of MS: if  $Z^*$ ,  $Z$ , and  $C^*$  are self-imposed rules, looking at them as if they were someone *else's* rules means that one is not following those rules but only scrutinizing them, which, as we have seen, removes their necessary status. With regard to  $C^*$ , that means that  $C^*$  cannot, in the third person standpoint, be regarded as necessary, but will always appear as a contingency, "the rule that *they* follow," a "habit," a physiological structure or "human nature." A rule that I am following is a necessity within the limits of my enterprise, but it is simply a contingency of my behavior to someone else. Accordingly, the Kantian notion of "necessity," from the third person standpoint, cannot even arise. Again, Quine and other "naturalized epistemologists" are correct, but only because they neglect the MS position altogether. One might add that Kant and Hegel, of course, would not accept this characterization of their position at all: they would agree with my analysis of  $Z^*$ ,  $Z$ ,  $C^*$  as rules, but then would object that, because the "enterprise" for which these function as rules is the whole of human experience, there can be no "outside" or third person point of view available to us. Accordingly, they would argue, one must treat these principles as one's own, and treating them as others' principles always presupposes their use in one's own case.

Kant argues that  $C^*$  principles function as rules, but he was not clear about the status of  $Z$ , describing  $C^*$  and  $Z$  alternatively as rules and as "structures," and so he was not willing to entertain the idea that  $C^*$  might be a matter of choice. Fichte then suggests that  $Z$  principles and consequently  $C^*$  principles are a matter of choice, that "the philosophy one chooses is the kind of man one is." Hegel takes up Fichte's suggestion, and the "forms of consciousness" of the *Phenomenology* are systems of self-imposed rules  $Z$ . Only  $Z^*$  is not a matter of choice.

The ultimate challenge to MS is scepticism, the

threat that  $Z^*$  might degenerate into solipsism. Traditional scepticism was a direct consequence of phenomenalism. Once one separates experience and physical object, even the love of God won't validly bring them together again. And so Kant rejects phenomenalism by insisting that our experience is necessarily experience of physical objects; and that must be so because of the rules by which *we* constitute our experience. And so, as Kant argues most explicitly in his "Refutation of Idealism," there is no place for the sceptic to get his wedge in between experience and object.

But the sceptic gets a second chance with Kant. There is no longer the vulnerable gap between experience and object, but there is a new gap, between the physical object as phenomenon and the object in itself. Now where did this gap come from? From an inconsistency in Kant's MSist analysis! Having completed his argument against phenomenalism and scepticism by pursuing more thoroughly than his predecessors the MS position, he now slips back into a naturalistic or third person standpoint and questions the MS position itself. Accordingly, he asks, in effect, "Even if it is necessary because of my a priori synthesis that I see objects in accordance with conditions which I impose (and by the a priori forms of intuition) is it not possible that the objects themselves do not obey these conditions?" But, the question is, from the MS position, illegitimate. Of course, one can question the MS position, as many recent philosophers have done, and reject it. But in that case, you reject it in the beginning, not when entering the climax of your epistemological deliberations. If Kant were consistent, he would have avoided this final slip out of the MS position. From the MS

position, the philosopher tries to prove the objectivity of what he knows; he cannot, and it does not even make sense for him, to ask what things are like apart from one's possible knowledge of them. Here Kant falters, and Fichte picks up the Cartesian banner—against the fading Kant's will—and carries it to Hegel. Hegel, a thorough-going MSist, never leaves the MS standpoint. Within it, he establishes, as Kant almost established, that from the MS standpoint, consistently carried through, there is no intelligible way to talk about, point to, grunt toward, or otherwise acknowledge the possibility of a world beyond our possible knowledge.

Now, stepping out of MS, its Hegelian conclusions might seem intolerable, much as waking up after a fabulous party the morning after: not only the particular festivities but the very idea of such merriment appears absurd. Stepping back into our everyday non-MS standpoint, it might appear that MS does end in the absurdity of the egocentric predicament after all. But this cannot be shown within the MS standpoint. Perhaps there are only two ways to follow consistently and honestly through the MS program in philosophy. There is Hume's schizoid way out, to leave his scepticism in his study and pick up his cue stick and his belief in the  $C^*$  belief in universal causation on the way to the billiards table; and there is Hegel's Kafkaesque pertinaciousness, maintaining the MS position by expanding its scope to the extent that, because there is no place else to go, there can be no question of a "predicament." There is ultimately "absolute truth," that MS, once taken up in earnest, can be consistently pursued beyond the point of possible refutation.



## IV. ON "EXISTENTIALISM", *EXISTENZ*-PHILOSOPHY AND PHILOSOPHICAL ANTHROPOLOGY

RICHARD SCHACHT

"EXISTENTIALISM" is by now a household word. Indeed, as long ago as 1947, Sartre could feel moved to complain (to a French audience, to be sure), "The word is now so loosely applied to so many things that it no longer means anything at all."<sup>1</sup> Still, this circumstance did not induce him to abandon the term as a convenient label for his views. And it is not difficult for almost anyone who has kept abreast of post-War literary and intellectual developments to state the basic tenets of "existentialism" in a rough and ready way; for they have been so extensively popularized in recent years that they are by now familiar to almost everyone who has any contact with the printed word.

The same is by no means true, on the other hand, of "*Existenz*-philosophy" and "philosophy anthropology"—expressions referring to two very important general European philosophical developments of the past half-century, which should be distinguished both from "existentialism" and from each other. My purpose in this paper is to draw these distinctions and to indicate something of the nature of each of these philosophical tendencies, focusing primarily upon those just mentioned. I shall begin, however, with a brief consideration of "existentialism" as it is commonly conceived and as it is characterized by various commentators, in order to provide my subsequent discussion with a familiar point of departure.

### I

As it is generally understood, the term "existentialism" refers to a certain comprehensive world-view, which is in striking contrast to what might be termed the traditional Judeo-Christian world-view, and also to most traditional philosophical interpretations of reality, whether rationalistic, idealistic, or materialistic; and an "existentialist" is one who subscribes to this understanding of the way things are. He is thought to maintain that there is no God

of any sort, and no absolutes in the realms of value and morality. There are only men and things. Men find themselves thrown into a cold and hostile world, condemned to an existence that is ultimately meaningless and fundamentally absurd, in which the only certainty is that of striving, suffering, and final oblivion.

Unlike mere things, the story continues, men are free in the most radical sense, the acknowledged limitations placed upon them by historical and social circumstances, heredity, mortality and the like notwithstanding; for they are always confronted by a range of alternatives, and within that range the choices they can and must make are neither psychologically determined nor governed by objectively valid normative principles. Men thus are what they make themselves to be through the choices they make—or, as the saying goes, their existence precedes their "essence" (i.e., their determinate character)—and so they are completely responsible for the lives they lead. This may be an exhilarating prospect; but it is also accompanied by the greatest anguish since choice in the absence of fixed and objective guidelines of any sort is not an easy affair, particularly when the shape of the only life one has is at stake. There are, of course, other men with whom one shares one's human lot of condemnation to absurd freedom; but no lessening of one's freedom and responsibility for one's choices is to be found in this fact. Indeed, while men share a common ontological constitution and a common human condition, they are radically isolated from each other, never being capable of achieving more than the most superficial sort of unity, and constantly struggling against each other except when temporarily making common cause in some struggle against still others.

In short, on the "existentialist" view (as this is usually understood), each person exists as a kind of monad, locked into his own subjectivity, and yet continuously confronted with objective realities—

<sup>1</sup> Jean-Paul Sartre, "Existentialism is a Humanism," *Existentialism from Dostoevsky to Sartre*, ed. by W. Kaufmann (New York, 1956), p. 289.

such as things, others, institutions, and his own "facticity"—with which he must deal without ever being able to achieve any real unity or even harmony with them. He may minimize his conflicts and his choices by making a general decision to follow the line of least resistance and take his cues from others; but he only deceives himself if he thinks that in this way he relieves himself of all responsibility for what he is. And one who does this further surrenders that possible authenticity or integrity which is the source of the only non-illusory sort of dignity attainable by man, in return for nothing more than a false sense of security and significance. He exists as a truly human being only if he recognizes his situation for what it is, in all of its insecurity, meaninglessness and absurdity; if he acknowledges his radical freedom and responsibility, resolutely accepting the inevitability of that anxiety which is inseparable from this acknowledgment; and if he charts a course of his own choosing, without any illusions of thereby rendering his existence less intrinsically absurd and more ultimately meaningful than it otherwise would be.

This, I believe, is more or less the sort of thing that "existentialism" is ordinarily taken—with considerable justification—to involve. And, if understood along these lines, the term is by no means a meaningless one; although to say this much is of course to say nothing about the plausibility and relative merit of the world-view in question. Problems arise, however, not so much from the fact that this characterization of the position is oversimplified and incomplete (as it admittedly is), as from the fact that the term "existentialist" is also commonly used to refer to a considerable number of rather diverse philosophers of European origin and recent vintage who would themselves take issue with this world view on a variety of counts. Indeed, the situation is further complicated by the fact that in the English-speaking world, "existentialism" and "recent Continental philosophy" are at least very widely regarded (even in philosophical circles) as virtually coextensive expressions. The latter notion is grossly inaccurate, as I shall shortly attempt to show by indicating something of the nature of another recent philosophical development in Europe, which cannot be assimilated to "existentialism" so conceived by any stretch of the imagination.

The former practice, on the other hand, of grouping together a large but *restricted* group of recent European philosophers under the rubric of "existentialism," presents a different sort of prob-

lem, at least where intelligent commentators who know what they are talking about are concerned. They realize that the position outlined above is by no means adhered to by all of those to whom the term "existentialism" has come to be applied (*viz.*, not only Sartre and Camus, but also Heidegger, Jaspers, Marcel, Kierkegaard, Nietzsche, and a number of others). Rather than attempting to restrict its application to those who do adhere to it (or to something close to it), however, they by and large have chosen to try to modify the description of what being an "existentialist" involves in such a way that the label may *accurately* be applied to all of those in question, however fortuitously it may initially have come to be applied to many of them.

There is something to be said for this move, given the difficulty of removing a brand once it has been burned in. It seems to me, however, that it would be greatly preferable to allow the term to retain its well-established and by no means illegitimate association with the general sort of world-view indicated above (thereby enabling it to retain its utility as a descriptive term with definite content); and to adopt an alternate way of referring to those philosophers to whom it is frequently applied, but whose substantive views differ significantly from those outlined, as well as to most of those who do subscribe to this world-view. This procedure might be just as questionable as the above-mentioned alternative to it, if it were the case that this would require an arbitrary terminological innovation. In fact, however, it would merely involve adopting a practice—and thereby making a very useful distinction—which has long been customary in Europe, and in Germany in particular. In the following section of this paper, I shall develop this suggestion, and shall elaborate upon my reasons for advancing it before going on to discuss several related matters.

## II

The term "existentialism" derives primarily from certain recent French philosophers—most notably Sartre (although Camus may also be mentioned in this connection). And it has become associated above all with a particular body of doctrine: namely, that which is set forth in Sartre's works written prior to his conversion to a form of Marxism in the early 1950s, and which is briefly outlined in the preceding section. This is not unusual. Terms ending in "-ism" generally are used to refer to some set of views. But precisely for this

reason, not many other philosophers—and very few who are not closely associated with Sartre—will admit to being “existentialists.” Heidegger and Jaspers, for example (the two most important German philosophers associated with the so-called “existentialist movement” in this century, as everyone knows), and the French philosopher Gabriel Marcel, are vehement in their rejection of the label. And they are quite justified in rejecting it, in view of the significant differences between their views and those of Sartre and Camus and their followers. To classify them under the rubric of “existentialism” is as inaccurate as it would be to term Kant’s moral philosophy a type of utilitarianism, simply because he is interested in ethics. Utilitarianism represents one particular philosophical position one can take in relation to various ethical questions; but ethics itself is simply a general problem area. The term refers only to the problems to be dealt with, and not to any particular answers to them. In the terms of this analogy, “existentialism” corresponds not to the latter (ethics), but rather to the former (utilitarianism).

Some commentators try to save “existentialism” as a general umbrella-term covering the positions not only of Sartre and Camus, but also of Heidegger and Jaspers, Kierkegaard and Nietzsche, and Marcel and others, by looking for a common element in all of their writings. Upon finding (or thinking they have found) some such thing, these commentators then proclaim that this common element is the essence of “existentialism”; and that a philosopher may be considered an “existentialist”—his objections notwithstanding—if this thematic common denominator is to be found in his works.

This is the approach of Blackham, for example, in his well known and widely used book *Six Existentialist Thinkers*; he justifies his selection of his six philosophers with whom he deals (viz., those mentioned above minus Camus) by maintaining that “together they develop the content of certain common themes.” What are these themes? Blackham says that the “existentialist movement” led by those he singles out “appears to be reaffirming in a modern idiom the protestant or the stoic form of individualism, which stands over against the empirical individualism of the Renaissance or of modern liberalism or of Epicurus as well as over against the universal system of Rome or of Moscow or of Plato.”<sup>2</sup> And at another point he says: “The

peculiarity of existentialism, then, is that it deals with the separation of man from himself and from the world.”<sup>3</sup>

Generalizations such as these, however, usually turn out to be only half true at best, and are more misleading than they are accurate and informative. They certainly do not provide satisfactory criteria for the inclusion of the writers popularly styled “existentialists,” and for the exclusion of others who most definitely are *not* so regarded. Blackham’s second statement above, for example, suggests that Hegel should be considered an “existentialist”; for no philosopher has been more concerned than Hegel—usually viewed as the exact opposite of an “existentialist”—with “the separation of man from himself and from his world.” And on the other hand, this statement does not apply very well to Heidegger; for he is much more concerned with our everyday *absorption* in the world of others and things and institutions than he is with any problems associated with separation from it.

Similar problems arise in connection with Blackham’s first suggestion. If advocacy of “the protestant or the stoic form of individualism” in some form or other is made the criterion of being an “existentialist,” then it would appear that Kant, Marx, and Bertrand Russell are as entitled to inclusion among “the existentialists” as are any of Blackham’s six. On the other hand, the use of this criterion would seem to render the inclusion of Nietzsche highly questionable, for reasons I shall indicate toward the end of this paper; and while Jaspers may be taken to satisfy it, it is a grossly one-sided and distorted characterization of his position, since he draws heavily not only on Nietzsche, but also on Hegel (as well as on Kierkegaard).

A different kind of attempt to find a common denominator is made by Walter Kaufmann. Kaufmann finds the essence of “existentialism,” not in any particular doctrine, but rather in the individualism of the various writers *themselves*—as thinkers and as human beings—and their critical attitude toward philosophical schools and systems generally. “Certainly,” says Kaufmann,

... existentialism is not a school of thought nor reducible to any set of tenets. The three writers who appear invariably on every list of “existentialists”—Jaspers, Heidegger, and Sartre—are not in agreement on essentials. Such alleged precursors as Pascal and

<sup>2</sup> H. J. Blackham, *Six Existentialist Thinkers* (New York, 1959), pp. v-vi.

<sup>3</sup> *Ibid.*, p. 151.

Kierkegaard differed from all three men by being dedicated Christians; and Pascal was a Catholic of sorts while Kierkegaard was a Protestant's Protestant. If, as is often done, Nietzsche and Dostoevsky are included in the fold, we must make room for an impassioned anti-Christian and an even more fanatical Greek-Orthodox Russian imperialist. By the time we consider adding Rilke, Kafka, and Camus, it becomes plain that one essential feature shared by all these men is their perfervid individualism.

The refusal to belong to any school of thought, the repudiation of the adequacy of any body of beliefs whatever, and especially of systems, and a marked dissatisfaction with traditional philosophy as superficial, academic, and remote from life—that is the heart of existentialism.<sup>4</sup>

Now, some of those Kaufmann mentions may have been “perfervid individualists”; but this is hardly true of two of the three “who appear invariably on every list,” viz., Heidegger and Jaspers. Heidegger may have become mildly eccentric in his later years; but both were professors, whose lives (at least for their times and circumstances) could hardly be said to be unconventional, and who never approached the individualism and anti-establishmentarianism of a Kierkegaard or a Nietzsche—or even of a Russell or a Wittgenstein. And the repudiation of systems and dissatisfaction with traditional philosophy of which Kaufmann speaks are common to many contemporary philosophers whom no one would think of terming “existentialists,” such as Wittgenstein, G. E. Moore, A. J. Ayer, and J. L. Austin. More importantly, it is hard to see why the general sort of orientation Kaufmann has in mind should be called “existentialism.” The only conceivable reason would seem to be that it suggests a desire on the part of those in question to make philosophy “relevant” to human existence. But this desire is far from being the special property of the men Kaufmann names, and has been shared (if such a desire can be meaningfully spoken of at all) by such diverse philosophers as Socrates, Hegel, and Marx. If the philosophers Kaufmann names have nothing more than this in common, there ceases to be any point in even grouping them together, let alone in speaking of something called “existentialism” in connection with what they wrote and thought.

In Germany, the terms *Existentialismus* and *Existential Philosophie* are used, but only in connection with the philosophy of Sartre and those whose views are closely associated with his. Another term is used to refer to the general philosophical problem area, which stands to *Existentialismus* as epistemology does to positivism, and ethics to utilitarianism. This term is *Existenz-philosophie*. Heidegger and Jaspers are not lumped together with Sartre as *Existentialisten* (“existentialists”); rather, they are all considered to be (at least part-time) *Existenzphilosophen*—philosophers who take human “*Existenz*” for their theme (or at least for one of their themes). No attempt is made to define *Existenzphilosophie* in terms of certain fundamental doctrines, or a pervasive basic mood (as Mary Warnock does<sup>5</sup>), or a particular attitude toward traditional philosophy, or a shared personality trait. Rather, it is defined in terms of its *subject matter*: a certain distinctive (purported) dimension of actual or possible human “being,” which I shall attempt to characterize in a general sort of way later in this paper. And concern with this subject matter is the criterion for determining whether someone deserves to be considered as *Existenzphilosoph*. To my way of thinking, it is extremely valuable to be able to distinguish between “existentialism”—as an “ism”—and “*Existenz-philosophy*” as a branch of philosophical inquiry, concerned with the analysis of human *Existenz*. The two are only too often confused in the English-speaking world; and this confusion easily leads to a misunderstanding of the general nature of an important part of European philosophy since the First World War.

“*Existenz-philosophy*” is an expression commonly associated primarily with Jaspers by English-speaking commentators; and “*Existenz*” is often treated as a term central to his analysis of human reality, but which is so distinctly *his* that reference to it is best restricted to discussions of his views. In fact, however, this very term is given an equally central place in Heidegger’s “hermeneutic of *Dasein*,” which he himself characterizes as an “analytic of *Existenz*.”<sup>6</sup> In one of his most categorical and definitive pronouncements concerning the nature of *Dasein* (his technical term for human reality), he asserts: “Das ‘Wesen’ des *Daseins* liegt

<sup>4</sup> Walter Kaufmann, “Existentialism from Dostoevsky to Sartre,” in the anthology edited by him and bearing the same title (see Note 1), pp. 11–12.

<sup>5</sup> Mary Warnock, *Existential Ethics* (London, 1967), e.g., p. 57.

<sup>6</sup> *Sein und Zeit* (Tübingen: ninth printing, 1960), p. 38; cf. *Being and Time*, tr. by J. Macquarrie and E. Robinson (New York, 1962), p. 62.

in seiner Existenz."<sup>7</sup> He further remarks, in this connection, that:

... when we choose to designate the being of this entity as 'Existenz,' this term does not and cannot have the ontological signification of the traditional term 'existentia'. . . . The term 'Existenz,' as a *Seinsbestimmung* [i.e., designation of (a way of) being], will be allotted solely to *Dasein*.<sup>8</sup>

And when he refers to aspects of the "*Existenzstruktur*" of *Dasein*, he terms them "*Existenzialien*" to emphasize their fundamentally different character from the "categories" applicable to "entities whose character is not that of *Dasein*."<sup>9</sup> In short, it is quite appropriate to speak of "*Existenz-philosophy*" in connection with Heidegger as well as Jaspers. And much the same sort of case can be made in the instance of Sartre, once one takes into account the language difference. To be sure, Sartre uses the same term (in French, as in English, "*existence*") in connection with both human reality and non-human entities. But he is at pains to point out that human "*existence*" is a very different affair from that of mere things, having the character not of determinate *being* but rather of the *transcendence* of determinate being (albeit limited by the latter, in the form of one's situation and "facticity"), and thus, in a manner of speaking, of "nothingness," and freedom. It is held to consist in a process through which one "encounters himself, surges up in the world," and comes to be "what he makes of himself"—this being Sartre's gloss on his dictum that "*existence precedes essence*,"<sup>10</sup> alternately expressed (in *Being and Nothingness*) in the proposition that "*consciousness*" is an entity "*whose existence posits its essence*."<sup>11</sup> (Or again, as he ultimately puts it in the latter work, it consists of "*a perpetual project of founding itself qua being and a perpetual failure of this project*."<sup>12</sup>) Thus, he writes,

... with man the relation of existence to essence is not comparable to what it is for things of the world. Human freedom precedes essence in man and makes it possible . . . What we call freedom is impossible to distinguish from the *being* of 'human reality.'<sup>13</sup>

The details of Sartre's analysis of the distinctive

character of human "*existence*" are not of concern here. The point is simply that his entire "*phenomenological ontology*" centers around his analysis of its distinctive character, and that he himself supplies the warrant for referring to his discussion of "*the being of 'human reality'*" as an investigation of human "*existence*" in a special sense of that term. While Sartre (writing in French) does not actually use the German term *Existenz*, therefore, the legitimacy of subsuming his analysis of human reality under the general rubric of "*Existenz-philosophy*" can hardly be denied.

In short, it seems to me that it would be pointless of anyone to insist that these expressions should be restricted to discussions of Jaspers, and avoided in discussions of Heidegger and Sartre (and others), merely because Jaspers is the only one of the three actually to refer to what he is doing as "*Existenz-philosophy*." It might be argued that it would be preferable to use more neutral language here, and simply to speak of "*existence*" and "*the philosophy of existence*." This, however, would be (and is) a dubious practice, since it would tend to obscure the very distinction—between a certain special (opaquely subjective) dimension of human reality and garden-variety non-human existence generally—which these and other such writers are concerned to draw and to clarify and explicate. As Heidegger suggests, in the passage cited above, a term to mark this distinction is precisely what is needed here; and I am simply proposing, as a designation of this general enterprise, an expression fixing upon the term which both he and Jaspers actually employ for this purpose (no slight to Sartre intended).

Actually, I believe that another distinction should be drawn here as well. An *Existenz-philosopher*, I have suggested, is simply one who is concerned with the analysis of *Existenz*. As such, he need make no special claims for *Existenz*, such as attributing to it greater significance than any other dimension of human life, or maintaining that human self-realization is exclusively a matter of its cultivation. Philosophers who concern themselves with the issue of man's nature, however, often feel obliged to take a position (either explicitly or implicitly) on this second question. Most of these

<sup>7</sup> *Ibid.*, p. 42; *Being and Time*, p. 67.

<sup>8</sup> *Ibid.*

<sup>9</sup> *Ibid.*, p. 44; *Being and Time*, p. 70.

<sup>10</sup> "Existentialism is a Humanism" (see Note 1), p. 290.

<sup>11</sup> *Being and Nothingness*, tr. by H. Barnes (New York, 1956), p. lxii.

<sup>12</sup> *Ibid.*, p. 620.

<sup>13</sup> *Ibid.*, p. 25.

historically associated with *Existenz*-philosophy—including Jaspers, Heidegger, and Sartre, as well as Kierkegaard—have answered it affirmatively. But this does not establish that there is no point in characterizing *Existenz*-philosophy formally, as I have done. It only shows that these writers were interested not merely in analyzing the nature of *Existenz*, but also in making further claims for it. And I should point out that one is not *eo ipso* an “existentialist” even if one does so. Perhaps another term is needed here, such as “*Existenz*ist”; and if I may be allowed to use this inelegant term, my point may be expressed by saying that “existentialists” as I have characterized them are only one tribe of “*Existenz*ists”; and that there are other such tribes of equal interest and significance, who generally regard the tribe of “existentialists” either with dismay or with hostility or with pity.

It should also be observed that the idea cannot be sustained that the *Existenz*-philosophy of Jaspers, Heidegger and Sartre (not to mention “existentialism”) is the only significant and interesting recent philosophical development in Europe, even if attention is confined to the works of these three men alone; for it would be misleading and inaccurate to characterize any of them simply as an *Existenz*-philosopher. And *Existenz*-philosophy would be as seriously misunderstood if it were to be identified with their philosophical efforts generally, as it is when it is identified with “existentialism” as characterized at the beginning of this paper. This may come as no news to those familiar with the whole range of their philosophical efforts, but it is worth emphasizing here.

Heidegger, for example, is primarily an ontologist, and is only secondarily (or preliminarily) an *Existenz*-philosopher. The basic question for him is not that of the nature of man, but rather is that of the nature of “Being” generally. His analysis of *Existenz* does not stand or fall with his reflections on “Being,” and may be considered quite independently of them; but he never tires of repeating, in *Being and Time* and elsewhere, that his analysis of human existence is a mere “preliminary analysis,” undertaken simply to prepare the way for an analysis of the nature of “Being” as such. And so, while he has never carried out this long-promised analysis of “Being” in any systematic way, most of his later writings are related more directly to his ontological interests than to the question of the nature of *Existenz* as such, and thus cannot be considered part of the literature of *Existenz*-philosophy.

Similarly, Jaspers’ interests are by no means

confined to the nature of *Existenz*. Originally a psychologist of great importance in Germany, his interest gradually shifted to many of the questions of traditional metaphysics. He neither raises nor answers these questions in traditional ways, and uses a vocabulary which tends to disguise the fact that he is dealing with them. Yet he expressly states that the three volumes of his *Philosophy* correspond to the traditional threefold division of metaphysics into speculation concerning the world, the soul, and God; only he finds it preferable to speak instead of “world-views,” *Existenz*, and “Transcendence.” Thus the second volume of this work, which contains the most complete statement of his *Existenz*-philosophy that he has published, is only one of three; and the other two—not to mention many of his subsequent writings—deal at length with themes which cannot be subsumed under the category of *Existenz*-philosophy.

As for Sartre: since the publication of *Being and Nothingness*, he has moved steadily in the direction of Marxism; and, in his more recent *Critique of Dialectical Reason*, he has (as I have already mentioned) developed a position which has more in common with the thought of the early Marx and various neo-Marxists than it does with Heidegger’s *Being and Time* and his own earlier work (in which he had drawn heavily upon the latter). He therefore cannot simply be regarded as an *Existenz*-philosopher and advocate of the “existentialism” which he developed in that earlier period. To be sure, a shift in the direction of a Marxist conception of man’s nature and human life does not preclude *Existenz*-philosophical concerns. Indeed, whereas Sartre may be said to have left his erstwhile “existentialism” behind him, setting it aside in favor of a rather different world-view, the same cannot be said with reference to his commitment to *Existenz*-philosophy. He has not abandoned it, but rather has broadened his approach to the understanding of human reality in a way that incorporates it and at the same time complements it with an objectively-oriented Marxian philosophical anthropology; and thus, in his own particular way, he has attempted an integration of these two rather different concerns (about which difference and integratability I shall have something to say shortly). But precisely to the extent that this is true, it follows that in his more recent work he deals extensively with matters falling outside the scope of *Existenz*-philosophy proper.

In connection with Sartre, a word is in order about the relation between *Existenz*-philosophy and

ethics. Sartre's particular version of the former, to which the term "existentialism" may justly be applied, and in terms of which "existentialism" is generally conceived, is often regarded as a type of ethical orientation. This is in no small measure due to Sartre himself; for it is in effect treated as such in his popular lecture entitled "Existentialism is a Humanism."<sup>14</sup> From what I have said about the nature of *Existenz*-philosophy, however, it should be clear that the philosophy of *Existenz* as such is a general program involving commitment to no particular set of doctrines, and therefore entailing no particular moral imperatives or axiological conclusions. There is only one way in which the question of such consequences can arise within the context of this program; and that is if some such consequence were in some way entailed by a particular analysis of *Existenz*. Sartre and Jaspers both seem to take this to be the case, whereas Heidegger does not (proclaiming his "authenticity-inauthenticity" distinction to be purely descriptive and to have no evaluative significance). It is an open question, however, whether any such consequences actually are entailed by the analyses Sartre and Jaspers give, and indeed whether an analysis of this sort can have such consequences even in principle.

### III

At this point I wish to draw attention to the existence of another recent European philosophical development; for I would not encourage—indeed, I want to attempt to correct—the mistaken idea that even *Existenz*-philosophy in this broader sense (let alone "existentialism") has completely dominated European philosophy, or constitutes the sole important development in European philosophy, in the past half-century. And here I do not simply have in mind the survival of Catholic philosophy and the revival of Marxist philosophy; although the former is taken much more seriously by secular European philosophers than it is by most of their Anglo-American counterparts, and the latter is of much greater genuine philosophical interest than most of what passes for Marxist philosophy originating in English. The existence and autonomy of both (in

relation to the tendencies I have been discussing) is generally known, even if the works and even the identity of the more important proponents of each are unfamiliar to most English-speaking readers. Nor do I have in mind the attempt to achieve a kind of synthesis of Marxism and Sartrean "existentialism," in which there is now a great deal of interest in both Eastern and Western Europe; although this development is not to be taken lightly by admirers of Sartrean "existentialism," since Sartre himself went over to such an "existentialist Marxism," giving it one of its earliest and weightiest expressions in his *Critique of Dialectical Reason*.<sup>15</sup> I do think it important to point out that Heidegger and Sartre have no monopoly on phenomenology, that phenomenology is by no means merely the methodology of "existentialism" or even of *Existenz*-philosophy, and that in fact the phenomenology of Husserl could hardly be further removed from either one in terms of focus, method, and content. However, I shall not go into this matter here, since I have done so at length elsewhere.<sup>16</sup> I would merely call these other developments to the reader's attention.

There is another philosophical movement, however, which is perhaps more important than any of these other developments, and which emerged contemporaneously with both phenomenology and *Existenz*-philosophy. In fact, the two works which marked the emergence of this movement, as Heidegger's *Being and Time* did that of *Existenz*-philosophy, appeared almost simultaneously with the latter. This movement is commonly referred to as "philosophical anthropology." Despite its importance in European philosophy (and its considerable intrinsic philosophical interest), it is virtually unknown in this country; although it has a good deal in common with what is called "philosophy of mind" (but which long since has had a much wider compass than this label would indicate) in the English-speaking world.<sup>17</sup> Its neglect on this side of the English Channel may be partly due to the fact that "existentialism" is much more sensational and conducive to literary exposition, and therefore has completely overshadowed it. Another reason, however, may be that "philosophical anthropolo-

<sup>14</sup> This lecture may be found, among other places, in Kaufmann's anthology cited above (note 1), pp. 287-311.

<sup>15</sup> *Critique de la raison dialectique* (Paris, 1960). This work has not yet been translated into English; although Sartre's long prefatory essay has been translated and published as a volume in its own right, under the title *Search for a Method*, tr. by H. Barnes (New York, 1963).

<sup>16</sup> See my "Husserlian and Heideggerian Phenomenology," *Philosophical Studies*, vol. 23 (1972), pp. 293-314.

<sup>17</sup> The only extended English-language study of the writings of those associated with it is Marjorie Grene's *Approaches to a Philosophical Biology* (New York, 1968). While far from comprehensive, this is a most useful book as an introduction to and analysis of some of the more important literature associated with this movement.



gy" draws heavily on human biology and the social sciences, and therefore strikes philosophers anxious to avoid overstepping the boundaries between philosophy and the sciences as a questionable commodity. European philosophers generally seem to be less concerned with the boundary than with the subject matter—which is, as its name taken quite literally indicates, the biological, social, cultural, and intellectual being, man.

Before proceeding to discuss it further and contrast it with *Existenz*-philosophy, I should perhaps indicate a few of the more important and influential men and books I am talking about. There is, first of all, Max Scheler, earlier a phenomenologist, whose book *Die Stellung des Menschen in Kosmos* (1928)<sup>18</sup> set out the program of a "philosophical anthropology" and, through the interest it aroused, provided the initial impetus toward the actualization of this program (although Scheler himself died before he could make any further contributions to it). Next, there is Helmuth Plessner, whose early book *Die Stufen des Organischen und der Mensch* (1928)<sup>19</sup> is a classic in the field, and who has more recently written a good, short introduction to it entitled *Conditio Humana*,<sup>20</sup> and also, several years ago (1970), a longer study actually entitled *Philosophische Anthropologie*. And finally, in the first rank at any rate, there is Arnold Gehlen, perhaps the most important of the three, whose early book *Der Mensch* (1940),<sup>21</sup> which focuses largely upon the biological nature of man and its behavioral consequences, is complemented by later writings which deal extensively with the essential functions of social and cultural institutions in human life.<sup>22</sup> (Unfortunately, due to the unclear standing of "philosophical anthropology" and its eclipse by other philosophical movements, almost none of the works of either of the latter writers have been translated into English.<sup>23</sup>) The "structuralism" of Levi-Strauss in France may also be considered at least a kind of near relation to this development. And much of the work of Ernst Cassirer, who emigrated to the United States from Germany after it had begun to achieve some recognition there, may be viewed as

yet another contribution to "philosophical anthropology." Indeed, in his late book *An Essay on Man*,<sup>24</sup> he expressly associates himself with this movement; and this book, together with his monumental *Philosophy of Symbolic Forms*,<sup>25</sup> constitutes the most important such contribution presently available in English.

At first glance, the program of "philosophical anthropology" and that of *Existenz*-philosophy would seem to be identical; for the one is concerned with the nature of man, and the other with human *Existenz*; and at the very least the two would appear necessarily to be closely connected. Indeed I would suggest that it is pointless to insist upon the incompatibility and hence ultimate irreconcilability of the two, as some of the adherents of each have done; for however different their starting points and methods may be, their conclusions will be suspect unless they can be integrated into a single, unified account of what it is to be a human being.

They have developed along different lines, however. This is at least in part because each has focused upon a different aspect of man's nature. The philosophical anthropologists have been concerned with what might be termed the "universal" in man—in other words, that which all men have in common, in virtue of their common biological nature and socio-cultural existence. They concern themselves with the general forms or patterns of human life, which can be discovered through the objective study of the human body, human behavior, and human institutions. Philosophical anthropology is by no means necessarily or even generally naturalistic or behavioristic; but its leading proponents have tended to characterize man's nature in terms which do not make essential reference to any irreducibly subjective dimension of human life or to the uniqueness of personal existence and experience. (I shall have more to say about it in the concluding section of this paper.)

It is precisely this sort of subjectivity and uniqueness, however, which has been the focus of the investigations of the *Existenz*-philosophers. When they speak of *Existenz* (or the same thing in other

<sup>18</sup> Bern and Munich: Franke Verlag, Sixth Edition, 1962. Published in English as *Man's Place in Nature*, tr. by Hans Meyerhoff (New York, 1961).

<sup>19</sup> Berlin, 1965.

<sup>20</sup> Pfullingen, 1964.

<sup>21</sup> Frankfurt a.M. and Bonn, 1966.

<sup>22</sup> E.g., *Urmensch und Spätkultur* (Bonn, 1958).

<sup>23</sup> The only exception, to my knowledge, is Plessner's *Laughing and Crying*, tr. by James Spencer Churchill and Marjorie Grene (Evanston, Ill., 1970).

<sup>24</sup> New Haven, 1944.

<sup>25</sup> In three volumes. Tr. by Ralph Manheim (New Haven, 1953, 1955, 1957).



terms), they are not referring to *all* dimensions or levels of human life, but rather to this one in particular. The term *Existenz* (or rather, its Danish equivalent) was first used in this special, narrower sense by Kierkegaard; and many *Existenz*-philosophers have followed him in this. These philosophers have tended to present *their* characterizations of what it is to be a human being along these lines, and so to take the position that man's essential nature or true humanity is to be conceived in terms of the uniqueness and subjectivity of the individual person. Philosophical anthropologists for the most part have not; and since the *Existenz*-philosophers have, there is a substantial disagreement between them (generally speaking) on this substantive issue.

But this difference is not a difference which precludes the possibility of any integration of the general programs of philosophical anthropology and *Existenz*-philosophy; for one does not *have* to take this position to be properly characterizable as an *Existenz*-philosopher (nor do philosophical anthropologists *have* to take some other, fundamentally different position to be philosophical anthropologists). Strictly speaking, an *Existenz*-philosopher is any philosopher who investigates the basic structures of human *Existenz* (or explores the meaningfulness of the notion), regardless of whether he conceives man's essence or true humanity in terms of it. Moreover, an interest in the nature of *Existenz* is by no means incompatible with the general program of philosophical anthropology. In fact, it would be quite reasonable to view *Existenz*-philosophy as a subdivision of a comprehensive philosophical anthropology, devoted to the analysis of one particular dimension of human "being" among others. (This, I might note, is precisely the position of a growing number of Western European philosophers today. They feel that the traditional antagonism between the two is pointless; that earlier *Existenz*-philosophers and philosophical anthropologists both offered one-sided and therefore inadequate accounts of what it is to be a human being; and that only something like an attempt to integrate the two is likely to yield anything approximating an adequate understanding of human reality.)

#### IV

I shall now turn to a brief consideration of the

question of what is to be understood by the term *Existenz*. Since the influence of Kierkegaard upon the development of this notion has been so great, I shall couch my initial remarks in terms of his approach to the matter as contrasted with the traditional philosophical treatment of the question of man's nature; and then I shall suggest a certain sort of basic problem subsequent philosophers have encountered as they have attempted to appropriate this notion from him, outside of the context in which he develops it, and some of the matters at issue between them.

Kierkegaard poses the question again and again: What does it mean to exist as a human being?<sup>26</sup> Initially, the question has a quite general meaning; the term "exist" has not yet been given a special sense. It is only as Kierkegaard undertakes to answer the question that the term comes to have a special, restricted meaning. In the writings of subsequent philosophers, the dimension of human life upon which he focused, and in terms of which he concluded that *truly* human "being" is to be conceived, has come to be associated with the term *Existenz*. The term can be employed in this connection, however, regardless of whether one agrees with Kierkegaard concerning its importance.

Thus where there was only one question for Kierkegaard—What does it mean to exist as a human being?—it is necessary for us to distinguish two quite independent questions: (1) What is the nature of that actual or possible (purported) dimension of human life to which Kierkegaard drew attention? and (2) Is this a privileged, uniquely important dimension, in terms of which alone man's essential nature or true humanity is to be conceived? (*Existenz*-philosophy, as I have characterized it, is the attempt to answer the first question. No affirmative answer to the second question is presupposed.)

Now, speaking very generally, it has been traditional in Western (not to mention Eastern) philosophy to conceive of man's nature in terms of characteristics or qualities which are such that what is unique and opaquely subjective about the individual cannot be expressed in terms of them. This implies (whether or not the implication is always explicitly acknowledged) that on these conceptions of man the subjectivity and uniqueness of the individual as such are something inessential. Consider, for example, the classical Aristotelian defini-

<sup>26</sup> See my "Kierkegaard on 'Truth is Subjectivity' and 'The Leap of Faith,'" *Canadian Journal of Philosophy*, vol. 2, pp. 297-313.

tion of man as a "rational animal." Here man is conceived in terms of two things: his reason and his animate nature. Reason is something entirely impersonal and universal. What is true in geometry, or physics, for example, or what is true as a matter of historical fact, is true for everyone if it is true at all, whether or not everyone actually recognizes it. In the realm of objective truth, idiosyncracies have no place, and must be set aside. It is this position which led to the conclusion, on the part of some classical philosophers, that while the mind may survive the death of the body, there would be no survival of personality in conjunction with it.

If one focuses instead upon Aristotle's other well-known definition of man, as a "social (or political) animal," the implications are similar. For, after all, social and political existence—existence and activity *qua* citizen or member of a socio-political order—is something just as impersonal and general in its own way as is intellectual activity. Here again, personal idiosyncracies and subjective feelings and dispositions are out of place and must be set aside; for in the realm of institutionally structured socio-political interaction, what counts is conformity to general norms and observance of established rules and conventions.

Similarly, man's fundamental animate nature is something universal among men. It is constituted essentially by certain basic physiological structures, associated with such general functions as nourishment, reproduction, movement, perception, etc. Characteristics which individuate one animate being from another of the same species were recognized by Aristotle to exist; but they were explicitly relegated to the realm of the inessential. (Not that one could characterize the uniqueness of the individual human being satisfactorily in such terms anyhow.)

Or consider Descartes' self-conception, set forth in his *Meditations*. He explicitly sets aside Aristotle's definition of man as a rational animal—and implicitly sets aside his definition of man as a socio-political animal as well—only to substitute the more simple definition of himself simply as a thinking being. ("I am not," he asserts, "this assemblage of members which is called the human body.") Thus, in the *Meditations*, we read the following:

Thought is an attribute that belongs to me; it *alone* is inseparable from my nature . . . I am therefore, to

speaking precisely, *only* a thinking being, that is to say, a mind, an understanding, or a reasoning being.<sup>27</sup>

All else is relegated to the category of the dubitable, and thus (for Descartes) to the realm of the inessential, where the nature of the self is concerned.<sup>28</sup> But as critics of Descartes have been fond of pointing out, there is nothing that identifies this thinking being as 'I'—that is, as a particular, unique individual self. If I am essentially nothing more than this, I am essentially nothing more than an instance of something quite general in nature.

Finally, consider the position taken by Hegel, who is of particular importance in connection with the emergence of *Existenz*-philosophy. Hegel does depart from the philosophical tradition to which I have been referring, at least to some extent. He argues that one essential trait of man is that of particularity; and he acknowledges the existence and ineliminability of a dimension of subjectivity in human life, in addition to its objective and impersonal dimensions. But while he thus does recognize that particularity, and its manifestations in conscious life in the form of "subjective will" and personality, cannot be neglected in forming an adequate conception of man's nature, he *does* consider this dimension of human existence to be of merely secondary importance in relation to that of "universality." Man, according to Hegel, is essentially "spirit"; and spirit is necessarily, and above all else, "universal." "Universality," in practical terms, means thought and action in accordance with universal principles—principles binding either upon the whole of a group of men (as in the case of laws and customs), or upon men generally, as thinking beings (as in the case of the sciences, logic, etc.)<sup>29</sup> And while Hegel recognizes that there is more to human life than such "universality," it should be observed that the Hegelian categories of particularity, personality, and subjectivity are conceived in such a way that they do not refer to anything opaquely inward or radically unique or intrinsically private about the individual and his experience. For Hegel denies the meaningfulness of speaking of an inwardness that cannot be outwardly manifested, and of a privateness that is not publicly accessible; and he has no principle in his system which even renders the radical uniqueness of the individual intelligible as a possibility.

Kierkegaard reacts strongly against this whole

<sup>27</sup> René Descartes, *Meditations on First Philosophy*, tr. not named (Indianapolis and New York, 1960), p. 26.

<sup>28</sup> *Ibid.*, p. 74.

<sup>29</sup> Cf. my "Hegel on Freedom," in *Hegel: A Collection of Critical Essays*, ed. A. MacIntyre (Garden City, 1972), pp. 289–328; and also my *Alienation* (Garden City, 1970), Ch. 2.

tradition, and against Hegel in particular. As he sees it (and as, following him, most *Existenz*-philosophers do as well), Hegel grasps neither the nature nor the importance of a kind of human individuality and subjectivity which is at least a possibility in human life, even if not a reality in most men's lives. Kierkegaard cannot abide the idea of the subordination of everything particular to the socially and intellectually "universal"; for on his understanding of man's nature, each person's uniqueness and subjectivity—his individual *soul*—are of paramount importance. He more than reverses Hegel: for him, in relation to this dimension of human existence, all others pale into virtual insignificance.

What Kierkegaard has in mind, however, is by no means mere empirical, objectively ascertainable differences between men. He emphatically does not conceive of the radical individuality of which he speaks in terms of physical characteristics; nor does he conceive of it in terms of habits and abilities and dispositions and preferences. He is ultimately thinking of the individual standing alone before a personal God who calls him to judgment and looks into the innermost recesses of his soul; completely stripped of the comfort and anonymity of the universal, and laid bare to his very depths—in which context observance of general customs, laws, and rules and intellectual attainments no longer have any positive significance. Thinking of oneself as related at every moment, by and as oneself alone, to a God concerned with what one has done with the particular life and freedom He has given to one, gives to all of one's thoughts and actions, and indeed to one's existence generally, a dimension of radical uniqueness, and at the same time bestows the greatest significance precisely upon what is unique about one. It is in terms of this possible dimension of human experience that Kierkegaard answers his question, "What does it mean to exist as a human being?"

To be sure, neither Heidegger nor Jaspers nor Sartre (nor, for that matter, virtually any other important philosopher associated with *Existenz*-philosophy, with the exception of Marcel) retains Kierkegaard's conception of God and of the individual's personal, radically individuating confrontation with Him. And this has meant that they have had to expend a good deal of effort attempting to come up with, and to establish the satisfactoriness of, some substitute—one, that is, which does not presuppose the validity of Kierkegaard's religious stance, but which nonetheless

performs the same function of radical individuation as does his "God-relationship." For Jaspers, who remains closer to Kierkegaard than either Heidegger or Sartre, the place of Kierkegaard's confrontation with God is taken by what Jaspers terms an encounter with "Transcendence" (a kind of demythologized version of the God of Judeo-Christianity and Kierkegaard); while for Heidegger, its place is taken by the individual's recognition of his "thrownness" into the world and of the inevitability and possible imminence of his own death; and for Sartre, its place is taken by the individual's awareness of his "condemnation to freedom" and of the absence of any objectively and universally valid guidelines which might direct him in the exercise of his freedom.

In all cases, however, an attempt is made to direct attention to the possibility and nature of a mode of existence in which a person is radically individuated (inwardly, though by no means necessarily in any objectively observable manner) in some such way. None of these writers deny that a great deal of what people are and do can and indeed must be understood in other terms; on the contrary. But as *Existenz*-philosophers their interest in these other actual and possible modes or dimensions of human existence is only secondary, and derives from the light which may be shed upon *Existenz* by a contrasting consideration of them. All are convinced that there is—or at any rate can be—more to human existence than can be expressed in terms of such impersonal notions as rationality, sociality, and animality; and also that there is more to it than anything which an observer could ever record, however closely he might attend to a person's appearance and behavior. And it is with this "more"—in virtue of which subjectivity is not strictly correlatable with objectivity, and individuality is not to be conceived merely as particularized universality or as the sum of various general forces converging on one point—that they, as *Existenz*-philosophers, are concerned.

This being the case, one may begin to wonder how it is that I could feel justified in my earlier contention that they differ among themselves too greatly to be held to have anything more in common than an interest in a certain subject-matter, other than the fact that those I have mentioned tend to be "*Existenz*ists" (in the sense indicated at the end of Sect. I above), and so to be more than merely noncommittal analysts of *Existenz*. To make this clearer, I shall attempt briefly to indicate some of the kinds of differences which divide *Existenz*-

philosophers—and thereby also *Existenzists*—at the level of their analyses of human *Existenz*. At the level of their general world-views, their differences could hardly be greater; for they take up a wide variety of positions on a broad range of the most basic philosophical issues: concerning, e.g., the existence and nature of some sort of divine transcendent being or principle; the possibility and nature of a universally binding scheme of moral principles and values; the nature and limits of human knowledge; the status of non-human objects; the kinds of interpersonal relationships of which people are capable; etc. However, differences at this level are relatively easily detected once one ceases to assume that there are none and goes looking for them; and they are only indirectly related to the issue of the nature of *Existenz*-philosophy. For this reason, I shall restrict myself to suggesting a few of the issues over which *Existenz*-philosophers as such often find themselves at odds with each other.

First, there is the matter to which I have already referred, of that which they take to perform the function of radically individuation, when the proper or appropriate relation to it is achieved. Few of them are at all in agreement about this, their proposals ranging from God to the prospect of one's own death. But this is by no means the only question over which they differ greatly. Another concerns the issue of the ways in which one's relationships with other men affect the realization of *Existenz*. Some contend that the realization of *Existenz* essentially involves engagement in a certain sort of interpersonal relationship; while others consider all such relationships to be entirely inconsequential to it; and still others regard all interpersonal relationships as a threat to it if not actually incompatible with it.

Another concerns the more fundamental issue of whether *Existenz* is something which *requires* to be realized at all. Some clearly hold that it does, and indeed that the realization of it is a very arduous process, which can never be accomplished once and for all; while others maintain that it is a permanent structure of human "being," to which we may remain oblivious, but which we ignore only at the cost of deceiving ourselves. Further, among the former, some contend that it is within our power to realize *Existenz* entirely on our own, while others hold that its realization is possible only through something like the extension of an act of grace from some source beyond ourselves, in addition to our own best efforts.

Another question concerns the issue of the nature of *Existenz* in relation to what is often summed up under the label of one's "facticity"—i.e., all of the "given" circumstances involved in one's being as one is in the situation in which one finds oneself. For some writers, *Existenz* is simply a matter of transcending one's facticity, in the sense that one is neither identical with it nor determined by it to any specific future course, but rather is always capable of taking up a variety of stances in relation to it. For others, on the other hand, *Existenz* essentially involves a certain sort of positive appropriation of one's facticity. This difference is related to (and may to some extent be clarified by) another, which concerns the issue of whether the personal uniqueness associated with *Existenz* in the life of a particular individual is completely acquired through, or in some sense and to some extent is given independently of, the choices he makes. The question, in other words, is that of whether one's *Existenz* is the sort of thing the specific "filling" of which is in a significant way *at stake* in the particular choices one makes, or whether it is unaffected by what these choices happen to be except insofar as they move one toward or away from the realization of it. Put somewhat differently: Is it to be conceived as a kind of pre-determined personal "essence," which one may or may not be true to (in which case it may be associated with the idea of "becoming what one is"), or simply as a fundamentally indeterminate general capacity of having to choose on one's own, in a kind of void in the sense that there is nothing more specific in the nature of one's self to be true to (in which case it may be associated with the idea of "existence preceding essence")?

*Existenz*-philosophers differ about this issue too, and about others as well, the complete enumeration and explication of which would require a good deal of space. It should not be necessary for me to go further into the matter, however, to make my point, which is simply that while they agree that there is a peculiar, opaquely subjective dimension of human life that philosophers previously have tended to overlook, the analyses of it they propose render it advisable to characterize their efforts collectively in terms of the area of their concern, rather than in terms of their general substantive agreements.

The fact that the latter are few indeed—in fact, virtually non-existent—may of course be seized upon by those sceptical of the soundness of the notion of *Existenz* and the enterprise of *Existenz*-philosophy, as providing ample grounds for concluding that their scepticism is well-founded. It

seems to me, however, that to leap from this fact to that conclusion would be rather hasty; for this state of affairs is by no means unique to the philosophy of *Existenz*. One has only to reflect upon the problem one would encounter upon undertaking to ascertain areas of substantive agreement among philosophers of stature who have been concerned with the philosophy of mind, or of morality, or of value, or of language or knowledge or the world of things and events, to conclude that a stronger case than this would have to be made to establish that the philosophy of *Existenz* deserves to remain beyond the pale of serious philosophical concern.

## V

I shall conclude with a few remarks about philosophical anthropology, in an attempt to make somewhat clearer both its nature and the distinctness of the "philosophical space" it occupies (particularly vis-a-vis *Existenz*-philosophy as characterized above). It seems to me to be best viewed as an outgrowth of the confluence of two philosophical traditions. One of them is that against which *Existenz*-philosophy emerged as a kind of reaction—viz., that the most notable post-Kantian representative of which is Hegel, to whose approach to and understanding of man's nature reference has already been made. The other was itself a kind of reaction against the latter—viz., what is known in Germany as *Lebensphilosophie* (the "philosophy of life," i.e., of the nature of life). This second tradition is often traced to Schopenhauer; and Bergson may also be mentioned in this connection. Most relevant for present purposes, however, is Nietzsche; and, partly for this reason, and partly because he is frequently thought to be among the precursors of "existentialism" and *Existenz*-philosophy, I shall introduce my remarks about philosophical anthropology by saying a few words about him.

In my view, it is a mistake to associate Nietzsche at all closely with *Existenz*-philosophy. To be sure, various *Existenz*-philosophers, such as Heidegger and Jaspers, have been greatly influenced by him; but then, they have also been greatly influenced by Kant and Hegel, which does not suffice to qualify Kant and Hegel as *Existenz*-philosophers. And that in Nietzsche which attracts them is something quite other than an analysis or even a conception of anything like *Existenz*. (In the case of Jaspers, it is

primarily Nietzsche's style of philosophizing; while in the case of Heidegger, it is above all his purported ambiguous relation to traditional metaphysics, as both its culminator and destroyer.)

The reason I take it to be a mistake to associate Nietzsche with the philosophy of *Existenz* is partly that I believe he would be highly unsympathetic to the very notion of *Existenz*, and would have regarded it merely as a late, dying echo of Christian "soul-atomism" (in his phrase), which continues to sound only because the full implications of "the death of God" have not yet been adequately recognized. But it is partly also because he is as completely committed to the idea of the correlativity of "inward" and "outward" as is Hegel; and further, because his preoccupation is with something altogether different from anything like *Existenz*.

Nietzsche's basic questions are of the following sort:<sup>30</sup> What is the fundamental nature of *life* in general? What is conducive and detrimental to it and its development? And, more specifically, in terms of what basic principle(s) are we to conceive of man's nature, and the quantitative and qualitative differences between men, as these are exhibited in their lives and actions? And what would the "enhancement" of "the type 'man'" (in his words) involve? He simply is not interested in the idea of some sort of opaque subjectivity and radical uniqueness which every individual might be supposed to have or to be capable of achieving, even if it were to be granted that he does not absolutely rule out the possibility of such a thing.

Nietzsche is closer to Kierkegaard in some ways than he is to Hegel (e.g., in his reservations concerning the possibility of men attaining anything that might be properly called Absolute Knowledge). However, he is closer to Schopenhauer (hardly an *Existenz*-philosopher) in many ways than he is to Kierkegaard—particularly in his general understanding of the nature of both reality and man. Schopenhauer broke with philosophical tradition when he took something non-rational—viz., "will"—to constitute man's essence, and indeed to express the nature of reality generally. Yet he followed in the tradition of employing a principle in characterizing man's nature in terms of which it is no more possible to express the individual's uniqueness than it was in the case of any of those who focused upon reason. Indeed, he even insisted upon

<sup>30</sup> Cf. my "Nietzsche and Nihilism," in *Nietzsche: A Collection of Critical Essays*, ed. by R. Solomon (Garden City, 1973), pp.58-82.

the ultimately illusory character of any *principium individuationis*, not only in nature but also among men. And Nietzsche followed him in this. Thus it is with good reason that, in Germany, Nietzsche is usually considered to be one of the first and most important representatives of so-called *Lebensphilosophie* rather than of *Existenz*-philosophy.

It by no means follows, however, that Nietzsche therefore is not as significant a figure philosophically as he is taken to be by those who associate him with the tradition of *Existenz*-philosophy. It only follows that his significance is at least largely to be conceived in other terms. If Kierkegaard provided the initial impetus for the emergence of *Existenz*-philosophy in the present century, Nietzsche may be said to have provided that for the subsequent emergence of philosophical anthropology, and also to have contributed a great deal to the discussion of those issues with which philosophical anthropologists in this century have been concerned. Both were convinced that the intellectual and rational activities of which men are capable have been greatly over-emphasized in most previous philosophical accounts of man's nature. Whereas Kierkegaard proposed displacing "mind" and "reason" in favor of the more subjective, personal, inward aspects of spiritual life, Nietzsche proposed displacing both in favor of the more objective manifestations of human life—and in particular, the human body and human conduct.

Like Hegel, but most emphatically unlike Kierkegaard and most subsequent *Existenz*-philosophers, Nietzsche contends that the only reliable guide to the inward is the outward, and that the key to what man may be said to *be* is what he may be observed to *do*. Hegel placed more emphasis upon social and "spiritual" expressions than did Nietzsche, who in turn placed more emphasis on the body than Hegel did; and this undoubtedly accounts for the fact that until relatively recently philosophical anthropologists tended to divide into two camps, one focusing more upon human biology, and the other more upon human culture. But in virtue of their subscription to this same basic methodological principle, their convergence was only a matter of time; and it is no longer easy to discern a greater affinity to the one or the other, even in the cases of writers like Gehlen and Plessner, whose biological orientation initially was quite pronounced.

Perhaps the best way to give some more concrete indication of the nature of philosophical anthropology in a very brief space is simply to indicate

some examples of (1) what various philosophical anthropologists have taken to be human phenomena having significance for the understanding of the kind of being we are; and a few examples also of (2) what concepts figure importantly in the thinking of some of the more important of them, and of (3) what some of the issues are which are a matter of disputation between them.

In the case of (1), a great deal of attention has been focused upon what might be termed biological differences between man and other zoological species, such as man's upright posture (and related characteristics), the unique development of certain parts of the human brain (and associated functions), and the underdeveloped state of the human organism at the time of birth; although in the more recent literature, attention has tended to center upon those distinctive human biological characteristics which are closely associated with elements of human socio-cultural life, by way of either facilitating or necessitating them. In addition, much discussion has been devoted to the consequences of the fact that while human beings may be said to have certain general needs and drives, they lack highly specified instincts and are capable of conditioning and adaptation to a degree far surpassing that characteristic of any other species. And finally (to mention only one more example), much has been made of a great variety of kinds of highly developed human behavioral phenomena, including not only the more obviously significant ones (e.g., tool-making, language-use, symbolization, the institutionalization at various levels of human interaction and need-satisfaction, etc.), but also such more commonly overlooked ones as aggression and war, tradition and habit, asceticism and self-destructiveness, luxuriation and boredom, and laughing and crying, as well as some which have recently become the objects of concern of quite a few English-speaking philosophers, e.g. emotion, punishment, "knowing how," and promising.

Examples of (2), as might be expected, are generally related in one way or another to the above. To be rendered comprehensible, let alone adequately explicated, a good deal of discussion of them would be required, which is impossible here. But they are much more than mere, vague catchwords; and the literature associated with them is well worth serious attention. They include instinct-reduction and unspecialization, world-openness and ex-centricity, sublimation and creativity, educability and acculturation, habit and tradition, historicity and sociality, symbolization and ab-

straction, and many others. (A brief but comprehensive and lucid summary and discussion of many of these notions and the related literature may be found in Michael Landmann's *Philosophische Anthropologie*.<sup>31</sup>)

As for (3), once again a few brief hints are all that can be given here. One general issue which has preoccupied and divided philosophical anthropologists is that of the extent to which various forms and aspects of human social, cultural, and ethical life are necessitated by or independent of human biological characteristics. Another is that of whether in the case of man "instinct-reduction" is so complete that human life is not even informed by certain very general "instinct-residues," or whether it is incomplete, with the consequence that there are significant limits to the plasticity of human nature. Another is that of the place and significance in human life—the relative necessity and/or dispensability—of such things as tradition, myth, conflict, and social participation. And another concerns the status of intellect and rational thought (in its various practical and theoretical forms) in the conduct of human affairs, in relation to the variety of non-cognitive forces and circumstances associated with human life as we must live it.

This list of examples, like those preceding it,

could be expanded at great length. Even from the little that has been said, however, several things should be clear. First, it should be evident that while some of the matters with which philosophical anthropologists concern themselves are closely connected with questions that require scientific investigation, their enterprise cannot be dismissed by philosophers as merely scientific rather than philosophical. Secondly, it should be clear that while some of these matters are of little current interest to most Anglo-American "philosophers of mind," others are at least reasonably close to some of those of concern to the latter; and that therefore it is not the case that philosophical anthropology is an enterprise of so strange a nature that no place for it is to be found within the realm of acknowledgedly legitimate philosophical endeavor. (Indeed, one would think that it would if anything be an enterprise much more congenial to both traditionally- and analytically-minded philosophers than *Existenz*-philosophy.) And finally, it should be obvious that philosophical anthropology represents a very different approach indeed to the analysis of human being than does *Existenz*-philosophy, the interest and importance of which are at least as great as those of the latter; even if, as was suggested earlier, the two do not exclude, but rather merely complement, each other.

University of Illinois (Urbana-Champaign)

Received May 24, 1973

<sup>31</sup> Berlin, 1964; esp. Parts III and IV.

## V. LINGUISTIC RULES AND SEMANTIC INTERPRETATION

SOUREN TEGHRARIAN

### I. INTRODUCTION

IT is commonly acknowledged that a speaker's linguistic competence cannot be explained in terms of his knowledge of grammar alone, that is in terms of phonology and syntax. Another facet of his linguistic competence, his semantic competence, needs to be taken into account. For modern linguistics the primary fact of language-use is that a speaker who is equipped with a limited supply of lexical items is able to form and understand an indefinite number of "correct" or "acceptable" linguistic expressions. The grammar of a language tries to explain this fact by articulating those rules which operate on grammatically simple elements (the phoneme in phonetics and the morpheme in the theory of syntax) to generate the grammatically complex phrases and sentences of the language. But the sentences thus generated by grammar still remain to be interpreted, i.e., understood. The task of semantics is to give an account of this interpretative ability exhibited in a speaker's understanding grammatically well-formed expressions.

For reasons too obvious to need stating, a philosopher of language is interested not in the grammatical but in the semantic aspects of the use of language. Encouraged by the impressive measure of success achieved by Chomsky and others in developing a theory of syntax in terms of explicitly formulated recursive rules to generate all and only the grammatically well-formed sentences of a language, some semanticists and philosophers of language have attempted to apply a similar treatment to the semantics of a natural language. The aim has been to arrive at a formally articulated theory of the "semantic structure" of language, in reference to which questions of interpretation and semantic acceptability can be answered objectively.

If this program could be carried out, it would have important consequences for the philosopher of an analytical persuasion who believes that philosophical questions are linguistic (i.e., semantic) questions in disguise. For if such a "structural semantics"<sup>1</sup> could be developed, philosophical investigations could be carried out in a systematic manner and questions settled with reference to the semantic structure of the language in which these questions are formulated.<sup>2</sup>

The characteristic contention in structural semantics is that sentential meanings are formally specifiable functions of the meanings of sentential constituents; that speakers of a language possess prior knowledge of the meanings of words, which meanings are further analyzable in the "lexicon" in terms of certain categories or sense-components, and with the help of some formally stated rules speakers are able to construct larger units of meaning (at the phrase or sentence level) out of these unanalyzable meaning-components of the words.

To examine the structural approach I shall refer to the semantic theory developed by J. Fodor and J. Katz (hitherto referred to as FK) which up to now remains the best worked out model of the structuralist claim. I shall present in outline the essentials of FK and in the rest of the paper attempt to show that such a model is inadequate for explaining the semantic properties of a natural language. I believe my criticisms will be applicable not only to the theory under consideration, but to any structural theory of semantics, since I shall be criticizing the basic presuppositions and those general features of FK that make it a structural theory. Similar criticisms, as I shall show, are applicable to any rule-based theory of meaning. I shall base my exposition of FK on Fodor and Katz's "The Structure of a Semantic Theory" (SST)<sup>3</sup>

<sup>1</sup> For a general survey of the problems in structural semantics as well as for an explanation of the term, see John Lyons, *Introduction to Theoretical Linguistics* (Cambridge, 1968), ch. 10.

<sup>2</sup> See Jerrold J. Katz, "The Philosophical Relevance of Linguistic Theory," in J. R. Searle (ed.), *The Philosophy of Language* (Oxford, 1971), pp. 101-120.

<sup>3</sup> Jerry A. Fodor and Jerrold J. Katz, "The Structure of a Semantic Theory," in J. A. Fodor and J. J. Katz (eds.), *The Structure of Language* (Englewood Cliffs, 1964), pp. 479-518.



and Katz's "The Semantic Component of a Linguistic Description" (SC),<sup>4</sup> since these works contain all the important elements of the theory.

## II. THE SEMANTIC THEORY OF FODOR AND KATZ (FK)

In SST Fodor and Katz first state their view of the different contributions grammatical theory and semantic theory make to the study of language. The aim is to show that certain types of linguistic knowledge cannot be explained by grammar alone. They base their contentions on the following: (1) sentences that have the same syntactic structure may differ in meaning (SST, p. 483), (2) synonymous sentences may have different syntactic structures (sentences in the active and passive voice<sup>5</sup>) (*Ibid.*), (3) syntactically unambiguous sentences may be semantically ambiguous (because of the presence of an ambiguous word) (SST, p. 485), (4) the context of a sentence may allow a word to have only one of its senses (*Ibid.*), (5) grammatically correct sentences may be semantically "anomalous" (*Ibid.*).

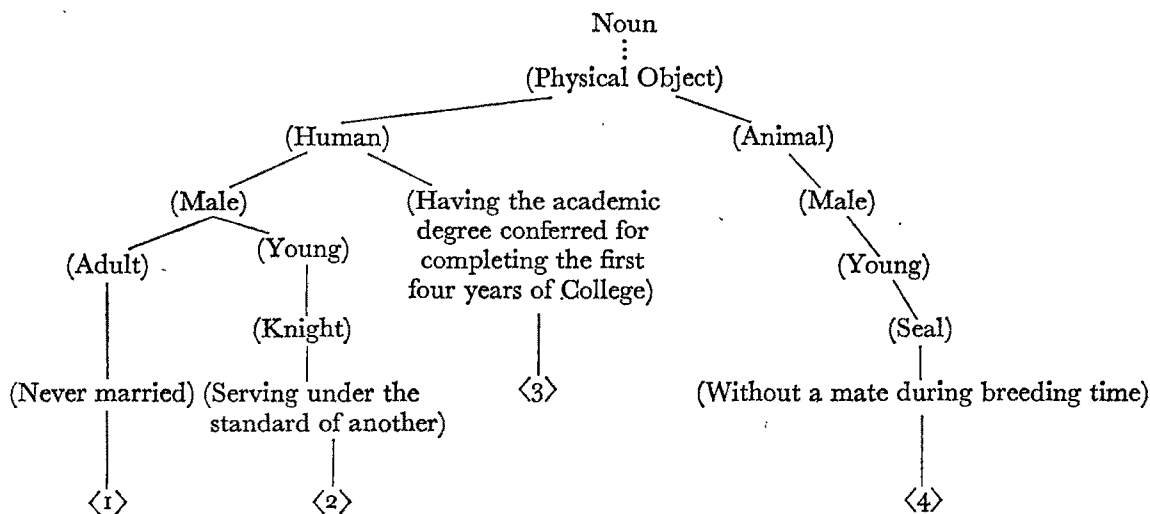
What sort of theory would account for these specifically semantic abilities? According to FK,

The basic fact that a semantic theory must explain is that a fluent speaker can determine the meaning of a

sentence in terms of the meanings of its constituent lexical items. To explain this fact, a semantic theory must contain two components: a dictionary of the lexical items of the language and a system of rules (which we shall call *projection rules*) which operate on full grammatical descriptions of sentences and on dictionary entries to produce semantic interpretations for every sentence of the language. (SST, p. 493).

A dictionary entry is represented graphically in the form of a tree-diagram. The entry for "bachelor" is given (SC, p. 185) as shown in the diagram below.

Each distinct path of the tree represents a distinct sense of the word and can thus be called a "reading." If a word (or "morpheme"—the distinction is not important for this paper) has  $n$  distinct paths it has  $n$  distinct readings or senses and is said to be  $n$ -ways ambiguous. Thus the word "bachelor" as illustrated in the diagram above is four-ways ambiguous. Furthermore, the tree-diagram embodies the contention that the meaning of the word "bachelor" is totally analyzable in terms of its sense-components, or in the FK terminology, its semantic markers (e.g., (Human), (Young), (Adult), etc.). The full dictionary entry of a word "consists of a set of finite sequences of symbols, each sequence consisting of an initial sequence of *syntactic markers*, followed by a sequence of *semantic*



<sup>4</sup> In Adrienne and Keith Lehrer (eds.), *Theory of Meaning* (Englewood Cliffs, 1970), pp. 176-198.

<sup>5</sup> This claim is surprising in view of the fact that modern grammatical theory is able to treat such sentences as *syntactic trans-*forms of each other. See Noam Chomsky, *Syntactic Structures* (Paris and New York, 1969), pp. 42-43 and 77-80.

markers, and finally, a *selection restriction*" (SC, p. 185). The latter specifies the *necessary and sufficient conditions* under which a particular reading of a lexical item can combine with others by the application of projection rules. Thus, the reading for one sense of the word "honest" carries the selection restriction (enclosed in angles)  $\langle(\text{Human}) \ \& \ (\text{Infant})\rangle$  which specifies that an adjectival occurrence of this word bears this sense if and only if the noun which this adjective qualifies contains the semantic marker (Human) but not (Infant). (SC, p. 189).

To give "a complete characterization of the notion of meaning" (SC, p. 191), we need, in addition to the dictionary, a set of projection rules which provide a *derived reading* for a phrase or sentence from the readings of its constituents. Thus, for the sentence "the child likes hard candy," we first obtain from the dictionary full information about each lexical item in terms of its semantic markers and selection restrictions. We then proceed with the syntactically simplest combination of morphemes—"hard candy"—and obtain a reading for it by the application of an appropriate projection rule. Then a reading for the phrase "likes hard candy" is derived by the application of another projection rule to combine the already derived reading with that of "likes," and so on until a reading for the whole sentence is obtained.

The following is an example of a projection rule (R1):

Given two readings associated with the nodes branching from the same node  $X$ , one of the form,

Sequence of morphemes  $m_1 \rightarrow$  syntactic  
markers of head  $\rightarrow (a_1) \rightarrow (a_2) \rightarrow \dots \rightarrow$   
 $(a_n) \langle 1 \rangle$

and the other of the form,

Sequence of morphemes  $m_2 \rightarrow$  syntactic  
markers of modifier  $\rightarrow (b_1) \rightarrow (b_2) \rightarrow \dots$   
 $\rightarrow (b_n) \langle 2 \rangle$

such that the string of the semantic markers of the head has a substring which satisfies  $\langle 2 \rangle$ , then there is a derived reading of the form,

Sequence of words  $m_2 + m_1 \rightarrow X \rightarrow (a_1) \rightarrow$   
 $(a_2) \rightarrow \dots \rightarrow (a_n) \rightarrow (b_1) \rightarrow (b_2) \rightarrow \dots \rightarrow$   
 $(b_n) \langle 1 \rangle$

Where any  $(b_i)$  is null just in case there is an  $(a_j)$  such that  $(b_i) = (a_j)$ .

This derived reading is assigned to the set of read-

ings associated with the node labelled  $X$  (SC, pp. 193-194).

This rule operates on constituents related by the modifier-head relation, that is, cases of adjective-noun, verb-adverb, adverb-adjective modifications, and so on. We would, for example, require a different projection rule for combining the readings of a verb and its object. Thus, the number of projection rules is determined by the number of distinct grammatical relations specified by syntax.

### III. SENTENTIAL MEANING AND SEMANTIC STRUCTURE: CRITIQUE OF FK

The system described in the foregoing is open to criticism on its three main aspects, which are of course interrelated: (1) the account of word-meaning in terms of semantic markers; (2) the account of sentential meaning in terms of projection rules; (3) the account of sentential meaningfulness in terms of the preservation of selection restrictions.

Obviously the FK model of word-meaning in terms of semantic markers is the foundation of the whole theory. And yet there are standard objections to this kind of approach which undermine, in my opinion, the whole scheme. It is simply a fact that, as Wittgenstein tirelessly emphasized, there are no necessary and sufficient conditions for the application of most words of a natural language, as selection restrictions in FK would require. I shall not dwell long on this point, which though important, has been amply demonstrated by Wittgenstein and his followers. It is therefore not surprising that Fodor and Katz rely for their analysis on a few carefully chosen words which have more or less precise meanings, such as "bachelor" (which is an example given again and again in their various publications<sup>6</sup>). But such words are the exception, not the rule. What semantic markers, for example, would the entry for the word "sweet" contain? The one which most readily comes to mind is (Taste). However, further subcategorization is needed in order to distinguish this entry from the entries for "bitter," "sour," etc. Yet it is hard to imagine what set of markers can do this job. Moreover, the adjective "sweet" can characterize tastes (e.g., the taste of an orange) as well as objects (the orange itself).

<sup>6</sup> See, in addition to the works cited above, J. J. Katz, "Analyticity and Contradiction in Natural Language," in *The Structure of Language*, *op. cit.*, p. 523; J. J. Katz, *The Philosophy of Language* (New York, 1966), p. 155; and J. J. Katz, *The Underlying Reality of Language and Its Philosophical Import* (New York, 1971), pp. 101-102.

Because of this, FK would characterize "sweet" as ambiguous, since in the one case the marker (Physical Object) is present and in the other not; and yet any native speaker would hesitate to maintain that the word has different senses in the two cases.

There are many questions about the meaning of the word "sweet" (as about most other words of a natural language) which cannot be answered categorically. Should its selection restriction carry the stipulation that it can combine only with nouns denoting edibles? (Can one speak of a sweet table?) What about the application of "sweet" to odors (such as the odor of a perfume)? Is it always possible to differentiate the various senses of a word so that on each occurrence of it in a sentence it is possible to select the appropriate reading? How finely should sense-differentiation be carried out without the danger of what Weinreich has called "infinite polysemy"? The way in which a word is understood varies with context, and yet it is not possible to say where one sense stops and another begins; and harder still, in the case of two distinct senses, to determine this difference in terms of a difference of semantic markers.

Consider such expressions as the following:

- 1a. To draw a picture of a house.
- 1b. To draw a picture of the economic situation.
- 1c. To draw a conclusion.
- 1d. To draw a distinction.
- 1e. To draw a boundary.
- 1f. To draw water from a well.
- 1g. To draw a pistol.
- 1h. To draw a lot.
- 1i. To draw a salary.

How many distinct senses do we have? If we consider the presence or absence of the marker (Physical Action) as a criterion we should put "draw" in 1a and 1f on the same branch of the tree-diagram (though they would still be distinguished by other markers) while 1b and 1c would belong to another main branch. On the other hand, the senses of "draw" in 1a and 1b, though separated by the marker (Physical Action) are more related than in 1a and 1f; and by this same criterion it would be difficult or impossible to classify 1h and 1e. Can one not draw a lot as well as a boundary (say between two states) through non-physical means? And if so, should we then say that the two

means of drawing a lot or a boundary indicate a difference of sense of "draw"? Such a conclusion seems unwarranted. Perhaps we should say that what decides whether there is a case of polysemy is whether a sentence such as "They drew the boundary along the river" is recognized as ambiguous by native speakers. But this won't do, because the aim of FK is to explain precisely this ability of native speakers to recognize such ambiguities in terms of the multiple senses of a word. There is no way in which a sentence can be ambiguous, according to FK, without containing at least one ambiguous (polysemous) lexical item. And there is no way in which a lexical item can be ambiguous except by having at least two distinct paths or readings characterized by a difference of at least one semantic marker.

The situation is not any better with projection rules. The latter are designed to explain how the meaning of a syntactically complex expression (a phrase or a sentence) is obtained from the appropriate senses of its words. A look at projection rule  $R_1$  however shows that it does nothing of the sort. All it does explain is that the "reading" of a complex  $W_1 + W_2$  somehow includes the readings (set of semantic markers) of its elements, when their respective selection restrictions are satisfied. No ordering of the semantic markers contributed by the two readings is provided for. In other words, the amalgamated sense of the complex  $W_1 + W_2$  is represented merely as the *unordered* set of markers of  $W_1$  and  $W_2$  when their respective selection restrictions  $R_1$  and  $R_2$  are not violated. But consider the example given by Weinreich:

- 2a. Cats chase mice.
- 2b. Mice chase cats.

Although these are syntactically ordered as  $a + (b + c)$  in 2a and  $(a + b) + c$  in 2b, the projection rules of FK would yield the same unordered set of markers for both, that is, they would yield the same derived reading for sentences which clearly differ in meaning. For similar reasons, Weinreich adds, "the theory is unable to mark the distinction between 'Three cats chased a mouse' and 'A cat chased three mice,' between '(bloody + red) + sunset' and 'bloody + (red + sunset)' and so on for an infinite number of crucial cases."<sup>8</sup> Hence, Weinreich correctly concludes, FK is unable to "deal adequately with the *content* of the readings of

<sup>7</sup> Uriel Weinreich, "Explorations in Semantic Theory," in D. D. Steinberg and L. A. Jacobovits (eds.), *Semantics* (Cambridge, 1971), p. 322.

<sup>8</sup> *Ibid.*, p. 321.

a sentence"<sup>9</sup> and that "the projection rules as formulated in FK *destroy the semantic structure* and reduce the words of a sentence to a heap."<sup>10</sup>

Granted that the FK projection rules are defective for the reasons given above, what other kind of rules would represent the "semantic structure" of a sentence in terms of its constituents? Weinreich seems to imply that whereas FK has failed to give a satisfactory answer to this, some other system could. But I think that the question itself presupposes a misconception about sentential meaning, and hence, not only FK but any structural approach to semantics is bound to fail.

What is at fault is the idea that the meaning of a sentence is the same sort of thing as the meaning of a word. On the basis of this supposition it has been maintained, as something almost axiomatic in contemporary semantic theory, that the meaning of a sentence is a "function" of the meanings of its words. FK has taken the explanation of this to be the main goal of semantic theory. The projection rules are functions which correlate the meanings of words to the meanings of sentences. But what sort of thing does the application of a projection rule produce? As we have seen, a projection rule is applied to sets of semantic markers to *produce another set of semantic markers*. But whereas the meaning of a word (and certain complexes such as a noun phrase) could, with some *prima facie* plausibility, be explained in terms of the latter's inclusion in certain categories (or semantic markers), the meaning of a sentence cannot. The meaning of a sentence is not exhausted by giving the semantic features of its words. A set of semantic features is not a sentence because it cannot do the job a sentence can do (i.e., perform a speech-act). It is somewhat plausible to treat a phrase such as "bloody sunset" as a single unit of meaning arising from the meanings of "bloody" and "sunset." Thus, a bloody sunset is anything which is both bloody and a sunset and the meaning of "bloody sunset" could be represented as the logical product of the class of features of "bloody" and those of "sunset." But the meaning of a *sentence* is not a set of features, properties, or characteristics. A sentence does not perform a purely referential function and so its meaning cannot be explained in terms of the set of features of its referent.

But isn't the meaning of a sentence a *function* of the meanings of its words? Obviously the operative

word here is "function." If we mean by this that the meaning of a sentence is determined by the meanings of its words, then the contention is no more than a truism—for a sentence is no more than words put together according to grammatical rules. But the word "function" is unfortunately understood in its logical or mathematical sense. Hence one looks for the functions or rules which when applied to word-meanings (in the way in which a mathematical function can be applied to the values of a set of variables) will yield sentence-meanings. This conception arises from the comparison of language with a mathematical system which is indeed the main motivation behind the structural approach to semantics.

In the case of a mathematical expression, its numerical value is a function of the numerical values of its variables. But here we have two crucial features that are absent in the case of sentences: (a) When a number is represented as a function of others (as when 15 is given as a function of 3 and 5) there exist alternative ways of identifying the number in question (e.g., as a function of 11 and 4). But when one says that the *meaning* of "The cat is on the mat" is a function of the meanings of its words, there is no independent way of arriving at this "meaning" other than through the self-same words (even "The mat is under the cat" won't do). If one asks for the numerical value of " $3 \times 5$ " the answer is 15. But if one asks for the meaning of "The cat is on the mat" the answer is not a "meaning" arrived at independently of these same words in the same arrangement (that is why perfect paraphrases do not exist). (b) When one quantity is said to be a function of another, as when the volume of a gas is said to be a function of its temperature, at least part of what is meant is that one quantity varies with the other. It is impossible to represent this contention except through the use of variables (a dependent and an independent variable). But when one says that the meaning of "The cat is on the mat" is a function of the meanings of its words, does one imply that if one varies the meanings of its words the meaning of the sentence will vary accordingly? Is a word such as "cat" a *variable* ranging over a domain of meanings? The only counterpart in linguistics of the idea of a function is found not in semantics but in syntax. In the syntax of a language one finds the counterpart of the mathematical notion of a variable in such

<sup>9</sup> *Ibid.*, p. 310.

<sup>10</sup> *Ibid.*, p. 321. His underlinings throughout.

abstractions as "subject," "direct object," "transitive verb," "count noun," etc., which range over certain domains of lexical elements, in such a way that different sentences are generated by substituting different lexical elements in the specified categories. But once a sentence is thus generated by syntax, its meaning is thereby fixed and subject only to contextual variations. There is no longer a question of making further substitutions of meaning for its words and thus obtaining different meanings for one and the same sentence.

A sentence is composed of words according to specifiable rules. But the meaning of a sentence is not composed of anything; rather, what we call the meaning of a sentence is what the sentence can be used to say or do in various contexts as an utterance. Sentences are tools with which speech-acts are performed. Because of the indefinite number of such acts, indefinitely many such tools are needed. Hence the need for sentences. To recognize a speech-act, we must recognize the tool, i.e., the sentence. In recognizing a sentence a speaker not only recognizes the words but also their composition. Up to this point a purely formal account is possible—the compositional patterns of sentences follow more or less rigid rules. But in understanding a sentence, in knowing what speech-act is or could be performed by its utterance, a speaker does not apply any further rules of composition. The type of knowledge that comes into play here is not a knowledge of rules but a knowledge of "extra-linguistic" situations and contexts, a knowledge of what other speakers say when, how, and why.

It is a mistake to think that just because grammatical rules do not constitute a complete account of linguistic knowledge, another set of rules is needed to complete the picture. I have so far shown that sentential meaning cannot be explained in terms of extra-grammatical (semantic) rules. However the existence of grammatically correct sentences which are nevertheless unacceptable as being "absurd" or "meaningless" is the main concern in contemporary semantic theory. I shall therefore examine the weaker view that even if semantic rules are not needed in order to obtain sentential meanings from word-meanings, they nevertheless are essential in order to account for sentential *meaningfulness*. According to this view, linguistic expressions which violate such rules (in FK, selection restrictions) are absurd or meaningless.

To begin with, many such expressions are, or can be considered as being, metaphors. Now to say that an expression is a metaphor is to say that it is meaningful, though perhaps "odd" when the words are "taken literally." Examples are abundant.

In Eliot's often-cited line

3a. The evening is spread out against the sky.

the adjective "spread out" is not used literally, unlike in the sentence

3b. The net is spread out against the wall.

How does a speaker understand or interpret 3a? In FK we find no answer. Perhaps the excuse for this might be that the theory is designed to deal with literal discourse only. In the light of this it may be viewed either as incomplete but expandable to deal with metaphorical discourse, or it may be defended as it stands on the basis that the semantic phenomena of a natural language are in any case confined to the literal uses of words.

This latter view is unacceptable. The semantics of a natural language deals with any non-grammatical competence of a speaker exhibited in his understanding or interpreting grammatically well-formed expressions. Fodor and Katz accept this when they claim "Linguistic description minus grammar equals semantics."<sup>11</sup> In fact, a speaker's ability to interpret metaphorical expressions is a demonstration of a highly sophisticated skill in the exploitation of the resources of his language. The belief that an account of metaphor can be conveniently left out of the theory of meaning is fostered by the erroneous view that such expressions are stylistic devices for saying what can be said literally. But it should be realized that finding a paraphrase is not any easier for a metaphor than it is for a literal expression. The meaning of 3a depends no less on the meanings of its constituent words than does the meaning of 3b. All one can say is that in using 3a a speaker is engaged in a different sort of linguistic activity (what Wittgenstein called "language-game") than in using 3b. Hence they have different sorts of meaning. But the two sorts of meaning are related because the metaphorical sense of a word is dependent upon its literal sense. It is a fundamental feature of language as a communicative device that meaning is "public," that one cannot use a word in any way he likes and hope to be understood, whether he intends to

<sup>11</sup> SST, p. 483.

be understood literally or metaphorically. Hence metaphorical interpretation is not an arbitrary process, but one which is tied down to existing usage. It is therefore incumbent upon a semantic theory to explain how a speaker with a knowledge of the literal senses of a word can understand its metaphorical sense in a sentence. But this is something no structural theory can do, for the essential contention of such a theory is that sentential meanings are well-defined functions of already established word-meanings found in the dictionary. All such a theory can do is to reject as uninterpretable sentences in which the use of a word violates a selection restriction.

Although the remarks above show the inadequacy of structuralism in dealing with non-literal meaning they nevertheless over-simplify the picture. First, it is not true that all expressions involving uncommon uses are metaphors. In fact, there are a great number of so-called figures of speech and metaphor is only one among them. But this is a minor point. Secondly, there exists no sharp distinction between "literally correct" and "literally deviant" expressions, a point which militates against the existence of selection restrictions which enable a speaker to decide, at all times, when the use of a word has gone beyond the limits of its sense. Rather, there is a gradually changing continuum at one end of which there are grammatically correct sentences which have standard interpretations and at the other end grammatically correct sentences which are capable of almost no interpretation except in very esoteric contexts (i.e., contexts in which analogies with familiar contexts of use are stretched beyond recognition). Metaphorical expressions in general fall in the middle of this spectrum. If the meaningful use of words was governed by extra-grammatical rules, the latter being the equivalent in natural language of formation rules in a logical or mathematical system, then there would indeed exist clear criteria for distinguishing meaningful from meaningless expressions. But such criteria do not exist, not even for a distinction between "literally meaningful" and "literally meaningless" expressions. Consider the following group of expressions.

- 4a. Colorful shirt.
- 4b. Colorful description.
- 4c. Colorful smell.
- 4d. Colorful competence.

It is clear that in 4a we have the most frequent and etymologically the most primitive sense of

"colorful," while in 4b we have a sense which is also standard but which is clearly derived from 4a. In general, it is rare to find words whose various senses are unrelated. It is another serious defect of FK that it fails to explain how a speaker who is in general familiar with the most basic sense of a word can nevertheless without difficulty understand other literal senses. It is safe to assume, for example, that at some point in the history of the English language 4b and similar expressions were considered metaphorical, that is, while 4b was recognized as meaningful (i.e., capable of being understood or interpreted) it nevertheless constituted a departure from the standard sense of "colorful." This means that speakers who were familiar only with the sense of "colorful" found in 4a were nevertheless able, on the basis of this sense, to understand 4b. The frequent use of 4b and similar expressions established a new sense, now listed in the dictionary. This illustrates how many of the semantic changes that take place in a language are brought about—something an FK type of approach is unable to explain. If the violation of selection restrictions resulted in nonsense, that is, if departures from established senses were uninterpretable, then it is hard to imagine how words could come to acquire new senses or extend their old ones. The structural model rests on the supposition that only those combinations of words are possible which preserve the already established meanings of their constituents. But if this were so, language would be a static system (like a mathematical system) and not, as it indeed is, something which is subject to continuous and dynamic change.

That there exists no sharp line of demarcation between "literally correct" and "literally deviant" expressions is illustrated by the somewhat uncertain status of 4c. Is "colorful" there being used literally or metaphorically? If one described a bouquet of various flowers as having a colorful smell (i.e., vivid, full of variety and interest) would he be using the word in the same sense as in 4b and hence literally (according to established usage), or would this be a deviation from its established sense? In an attempt to answer this question we do not, as a matter of fact, appeal to rules. What we appeal to instead are analogies of various sorts. On the one hand, smells can be as vivid and as interesting as anything else, but on the other hand, while "colorful" as applied to objects (4a), descriptions, parades, etc. (4b), has visual referents, this is obviously not the case with smells. There is a strong analogy, but there also seems to be a shift from visual to an

olfactory referent. Many metaphors work by such shifts. So the answer to the question whether 4c has a literal or a metaphorical meaning depends on the crucial point whether the already standard sense of "colorful" in 4b applies *only* to visual referents. But how can we determine this? The dictionary is no help here, and the most fluent speakers of English will disagree among themselves.

Earlier, I suggested that grammatically correct sentences form a spectrum with varying degrees of intelligibility. Totally meaningless or completely uninterpretable combinations of words can be found only among ungrammatical sentences (or non-sentences) such as "rapidly blue don't is pen." However, some grammatically correct sentences such as Russell's "quadruplicity drinks procrastination" at the extreme end of the spectrum can come close to ungrammatical sentences in their unintelligibility.

That there are indeed expressions whose status is uncertain as far as the meaningful/meaningless distinction is concerned, is illustrated by 4d. While it could be a result of a misuse of "colorful" or "competence," it is nevertheless not impossible to discover contexts in which its use, though not quite natural or stylistically sound, would have *some* point and in which *some* analogies with more familiar contexts (such as 4b) would be preserved. Thus in the expression "He played the piano with colorful competence," the use of the adjective could be to contrast the performance with one which, though faultless and competent, is drab, uninteresting, or unexciting. What makes 4d less intelligible than 4c is the fact that grounds for comparison with expressions like 4b, in terms of which both are interpreted, are less available for 4d than for 4c. In 4c while the shift is from a visual to a non-visual, but still a sensory referent, in 4d the referent is not even sensory any more. Thus, from 4a to 4d we have a gradual loosening of analogies to the point where in the latter the analogy is so weak as to make the expression almost meaningless. An account of meaningfulness in terms of rules cannot explain the facts mentioned above, while the approach I have been suggesting explains why meaningfulness admits of degree: it is because analogy does likewise.

My arguments show that the intelligibility of the use of a word in an expression is (1) subject to degree, and (2) determined by the existence and strength of analogies with familiar contexts. One

could still maintain that there are semantic rules, not of the sort found in FK, but ones which are loose, flexible, and open-ended. Although this claim is not strictly refutable, I think it is superfluous. I agree with Paul Ziff that "Speaking and understanding a natural language has nothing to do with rules."<sup>12</sup> This is true at least to the extent that speakers of a language do not appeal to rules when they contemplate the meaning or even the meaningfulness of a grammatically well-formed expression. But of course, the contention usually made is that such rules are known and applied *tacitly*. That semantic rules exist is said to follow from regularities in the use of words, and from correct and incorrect usage.

That there are indeed regularities in the use of words (over and above those captured by grammatical rules) is a fact which cannot be denied; and that the knowledge of such regularities is a necessary condition of a speaker's ability to communicate successfully cannot be denied either. But it should be realized that such regularities in the case of most words are characterized by the existence of similar, rather than identical, elements in familiar contexts of use. This is the reason why if there were any semantic rules, they would have to be loosely formulated. But there is another problem. A rule is either obeyed or violated and the resulting expression is either correct or incorrect. Such a situation can be described in grammar more naturally than in semantics. Semantics deals with the intelligible use of words, and intelligibility, as I have shown, admits of degree and varies with context, while obeying a rule and the corresponding notion of correctness do not. This is why well-formedness is an ill-conceived model for semantics.

A semantic rule as an expression of some regularity would enable us to spot deviations from the latter. But whether the use of a word constitutes a deviation from a regularity established by speakers is, more often than theorists seem to think, impossible to tell. Hence semantic rules would have to be open-ended. But even where there is a clear deviation from an established regularity, the resulting expression may still be intelligible. Speakers of a language not only share a knowledge of typical contexts of use but also a knowledge of different sorts of linguistic activities, including the activities of deviating from regularities of use. And since this is so, the notion of correct use (in accordance with

<sup>12</sup> *Semantic Analysis* (Ithaca, 1960), p. 34.

a rule or regularity) cannot be coextensive with that meaningful or intelligible use.

Deviations from grammatical regularities result in ill-formed expressions, while deviations from semantic regularities do not necessarily result in nonsensical utterances. While in order for a speaker to manage to *say* anything at all it is necessary for him (by and large; slight distortions of syntax are not a problem) to construct a well-formed sentence, whether what he thus manages to say (e.g., "quadruplicity drinks procrastination") makes sense, and to what degree, is not the same as whether his use of words preserves exactly those conditions under which other speakers use the words. Of course, there *has* to be a connection (between typical contexts of use and the new one), otherwise it would be impossible to interpret what he says; and the relative strength of this connection determines the type and degree of meaning it has,

ranging from literal to metaphorical, and ultimately to lack of meaning.

What the rule approach to semantics fails to take into account is the fact that our everyday use of language is to some extent a creative process and not a mechanical one, as our discussion of semantic interpretation has shown. The conditions for the use of a word that become standard are the basic data upon which communication is based. These data have not only linguistic, but ultimately social and philosophical significance as well. This much has been acknowledged. What is however not generally acknowledged is that deviations from these standard conditions may also have philosophical significance, in the sense that they may be expressions of philosophical *insight*—a result of discovering similarities and connections between phenomena, hitherto unnoticed and not yet incorporated into the language.

• Fairleigh Dickinson University

Received August 13, 1973



## VI. SIGHT AND LIGHT

VIRGIL C. ALDRICH

THE funsters in philosophy are those who like to stagger the imagination with notions of what is logically or theoretically possible, stating this in a way that suggests not only that these suppositions are contradiction-free, but that they might be actually true of some existing state of affairs not only elsewhere—say on Mars—but even right here and now. What gives the green light for such procedure is the assumption that, given a cluster of terms naturally affiliated in plain talk, what each of them “means” is nevertheless logically independent of the others, such that combinations of them in declarative sentences express what is only contingently true. Thus, for example, what “see” means is said never to be *conceptually* tied with what “eye” or “light” means. So, it is theoretically (logically) possible for someone to see something without light or eyes. The philosophical funsters, then, are those who tend to think of each concept as having its own essence, which logical autonomy allows it to be meaningful even when thus construed in abstraction from natural affiliations with other concepts, as it is when applied to something to which the affiliates can *not* be significantly applied. Plato and Descartes are magnificent exemplars of this style of thinking, and they have followers who currently carry on as if Aristotle, Wittgenstein, and Austin had never lived. (Kant has a leg in both camps; so has Strawson.)

The aim of this short essay is modest. It does *not* attempt to show what, in general, is wrong (if anything) with funsterism in philosophy. I am convinced that a good half of the story of philosophy will continue to be about great philosophical funsters, since philosophy is the tilt, the tension, between them and the workers in it. (After Zeno of Elea and Plato, there had to be an Aristotle; after Hume, a Kant; after Russell, a Wittgenstein.) Anyway, it has become fatuous to hurl such blunderbus charges as “essentialist fallacy,” “abstractionism,” “intellectualism,” and the like at the funsters. Instead, what follows is a set of staccato considerations—questions and remarks—that may help us on to some fresh reconsiderations of the concepts in the cluster containing “see” and “light.” These reflections will, to be sure, make

one feel a tug in a direction that the funsters will be least inclined to move. But they are not here going to be pushed in that direction primarily by *arguments*. Be that as it may, my Parthian shot about philosophical funsters is that the way they argue affects one as the poem “Jabberwocky” affected Alice. In her words: “Somehow it seems to fill my head with ideas—only I don’t know exactly what they are.”

Before beginning the specific treatment of funsterish theses about sight and light in the next (main) section, let me state the *general* position that perplexes philosophy by not making enough of the role of the body in perception. It is spawned under the spell of a phenomenalist or Cartesian picture. According to the misconception this spawns, at one end of a continuum we have a big, blooming confusion of raw, subjective sensations “in the mind,” problematically related to an objective, external order of things. This indeterminate “sensory manifold” is gradually objectified (organized) by conceptualization or categorization into the objects of ordinary perceptual experience. But the mind cannot come to rest in what it thus perceives, because these homely things are only its first-order constructs, theory-laden with an implicit theory that the mind must go on to make explicit, in a “conceptual system.” Thus is simply seeing something pictured as a simple-minded bit of *thinking* that something is the case. So the mind moves on to more sophisticated theorizing, consummated in the notion that what and where the thing actually is, its “nature,” is in principle completely out of sight. Thus is perception made to look like half-baked (unfinished, unconsummated) conception. *Percepta* are volatilized into *cognoscenda*. In current philosophy of language, this idea appears as the notion that all (categorematic) terms are “theoretical.”

What is wrong with this picture—it *is* a picture, not a concept—is that it presents only a single development, a maturing only of *thinking*, a development only of the mind toward reliable beliefs formulated in a theory. As if the only reliable vision of things is theoretical. But “vision” is a bipolar concept. At one end it is simply seeing things; at the

other it is consummated thought. What is neglected by the above wrong picture that spell-bound even Plato is the fact that people mature not only into thinkers but also into seers in the sense of people with good eyesight, or with acute vision tested in ways that have nothing to do with conceptual capacity. This is a growth distinct from, but concurrent with, the development of vision in the theoretical sense (seeing that such-and-such is the case.) Moreover, it is presupposed by the latter, the simply seeing being that which presents the "one world" (Dretske) that theorists render in this or that conceptual system.

Now I turn back to the specific concern about the logical conditions for the applicability of "simply see."

## II

The stage is set for a good part of what follows by Dretske's remark<sup>1</sup> that anyone who could correctly report the colors and color changes of something on the far side of a "massive" wall—without any mirroring devices in the offing—would still be said to "see" it, though we wouldn't know how he did it. Just as, if someone jumped 12 feet straight up, we would not say that he didn't jump. We would just not know how he did it, things being the way they now are. So, it is not a part of what it *means* to "see" that the view of the object be unobstructed by opaque things in the line of sight. Logically speaking, it is thus possible to see through a foot-thick slab of lead which stops even cosmic rays (my example).

To deal with this issue, we focus attention on the cluster of concepts containing "sight" and "light." The other relevant members of the set are "field of vision," "color," "eyes," "direction the eyes are turned," "looking at," "line of sight," "visual impression." The sense of "see" here under consideration is clearly connected with "good eyesight," not with the notion of "see" that is used, say, after an explanation. ("Now I see what you mean"; "I see that \_\_\_\_".)

1. What is the criterion for someone's (*S*) seeing something (*x*)? Is it, in *any* circumstances, *just* that he can tell, say, its color and shape, and report changes in these as they occur?

2. Suppose it is Descartes's piece of wax undergoing temperature—and therefore color and shape—changes. *S* is holding it in his hand, looking at it

in the light, making the above reports about shapes and colors. How can he tell is here readily answered by: *S* sees *x*. In these circumstances, his telling is clearly a criterion for his seeing. Had there been a question about his visual capacity—is he blind or color-blind?—his performance in that set-up would decide the matter. He can see, and tell by seeing.

3. Suppose now he continues correctly to describe these changes in the wax when it is placed on the far side of a massive opaque wall. Do we now say he sees it, and tells by seeing? Well, is he still looking at it, or are his eyes (say) downcast and half-closed as he makes his reports? If this is the case, his being able to report color and shape changes is not so clearly a criterion of his *seeing* the piece of wax. Here, one is naturally disinclined to say that he tells by seeing, and looks for an extraordinary non-visual rapport that *S* must have with *x*, on the strength of which he can report colors and shapes without even looking. And if looking is unnecessary, eyes seem to become dispensable also, or at least *S* need not engage them in the awareness of shape and color in this strange case.

4. Suppose, though, that we do not dispense with eyes and looking so fast. We can imagine *S* to have to "look" at *x* on the far side of the opaque wall, or at least to direct his eyes at it, to keep it in his "line of sight," if he is to continue making the shape and color reports. What then do we say about "seeing"? Does he *see x*?

First, there would be problems about what "looking" and "line of sight" mean in such a case. By hypothesis, the wall is "opaque," and that ordinarily entails something that cannot be "seen through." "Opaque" has a tap-root connection with "dark"; anything opaque between the would-be seer and the object leaves the latter "in the dark" for him. "Opaque" in this primitive use, must be distinguished from the theoretical concept of impenetrability by light rays from the object, the concept that plays the major role in current causal theories of vision. In its primitive (and still viable) use, "opaque" is applied to what terminates the *line of sight*—an obstacle which must be taken out of the way if *x* is to be seen.

But, suppose that "line of sight" is more optically determined here by the direction of the eyes, their convergence and accommodation, etc., and all this is just as it would be were *S* to be seeing *x* with no opaque wall between. What then do we say about

<sup>1</sup> Fred Dretske, *Seeing and Knowing* (Chicago, 1969), pp. 50-51.

*S* seeing  $x$ , as he correctly reports changing colors of the wax on the far side of the wall?

Notice, first, how tricky the confirmation would be of the ocular performance being exactly as it is in the normal, wall-unobstructed case. But, never mind. It is sufficient, in the funsterish game we are now playing, to "imagine" the ocular adjustment being pretty much as usual. Then, with what we know about radiation, we would be likely to suggest waves different from those that ordinarily are necessary conditions of sight. The trouble with this suggestion is that we already know of waves with greater penetrating power than light waves, but these have no connection with colors or color vision. "Maybe *S* sees the yellow of the wax with cosmic rays" is a conceptual howler. Yellow already is another wave-frequency. (Be careful about saying that this is a "contingent fact.")

5. But the game we have started to play, in relation to  $x$ 's color "seen" through the opaque wall, makes all these quasi-empirical considerations irrelevant. The point is, rather, that we can *imagine* *S* making the correct color reports, with or without the thing being illuminated and in sight, in the ordinary sense. The funster's point is that, given this, we would say that somehow he "sees" the wax though we wouldn't know how.

6. Next, suppose there is a hole in the wall through which *S* puts his arm, and can make the color reports only with his finger tips on the piece of wax. Would we then say he sees it, or that he can tell by seeing? Remember, we began by treating *S*'s ability to report  $x$ 's colors as a criterion for his seeing  $x$ .

But, in this new case, one is inclined to say that *S* tells colors not by looking and seeing but by touching and feeling. Or, if one hangs on to the notion of seeing here, thanks to its natural connection with the concept of color-awareness, then one is forced to introduce the odd notion of seeing things with one's fingers. Fingers do for eyes, in this case.

7. Again, empirical reminders of what has actually been achieved in this direction (the Russian and other experiments) are irrelevant. The case has been made to rest simply on what can be imagined, and who cannot imagine even a blind *S* telling colors by touch? (Just stretch Helen Keller's case a trifle, if you want to preserve the pseudo-empirical complexion of the investigation.)

8. Or by listening to and hearing the colors? Or just by sniffing  $x$  with one's nose? Or tasting it? Of course, one may want to delete "hearing,"

"smelling," and "tasting" in these odd cases—since it is color that is revealed—saying instead that *S* sees  $x$ 's colors with his ears, or nose, or with tongue and taste-buds. What *would* one say in these cases?

9. One thing is clear: the notion that being able to tell colors is intrinsically or *by itself* a criterion for seeing them is here put under a terrific strain—without at all tarnishing the notion that it is the criterion in normal circumstances. And not only is this critical notion staggered by removing, say, just light from the visual situation, but the other members of our cluster of concepts also begin to curdle. "Looking at," and its connection with "eyes" and "line of sight," blur as this opening wedge for conceptual confusion is inserted, namely, the notion that  $x$ 's being in the light is not, logically, a necessary condition for its colors being visible.

10. But, now that the wedge is in, let us watch what else flows in, to consummate the curdling of the concepts in our cluster.

11. Suppose *S*'s arms are amputated. He has no fingers to touch things with. But we imagine him able to report the cool or the warm—"temperature" in that rudimentary sense—of anything he simply sees. Simply by looking at the piece of wax he can report its temperature—say, it's being warmer than it was before; and *so for anything else he simply sees*.

Now, ordinarily, just as the criterion for *S*'s seeing something—or not being blind—is his ability to tell the colors, so the criterion for his feeling its temperature—or not being anaesthetized—is his ability to report the cool or the warm. In the strange visual case, we (Dretske) hung on to the notion of seeing color without ocular looking. Shall we hang on here to the notion of feeling temperature without touching or any bodily contact? And does this require us to say that, in this odd case, one is doing with his eyes exactly what one ordinarily does with his fingers?

12. These considerations have, in effect, brought the massive wall into this new picture. What if one could tell without touching (or looking), the temperature of the (say) cool wax on the far side of the wall? Does he tell by *feeling* it without touching, as he *sees* color without ocular looking in the other strange case? (Don't let the notion of *radiant* heat confuse this issue; it is confused enough without it.)

13. What has been coming to notice so far suggests a chain of concepts linked by logical relations that are hard to characterize. Their logic seems

"informal." At one end—call it the "objective" end of the series—is the colored thing. It is natural to think of color (in the primitive sense) as the quality of things that is revealed only by sight. So we naturally treat being able to tell color as a criterion for seeing. To be able to tell the color here and now is to "see" the thing. But this does not amount quite to a definition either of "color" or "see"—though the linkage is logical. Color is what you tell by seeing—though there would be the color without the seeing, so the seeing is not "of the essence." Then comes "light" as a necessary condition (not yet causally construed) of color vision. One can't see colors in the dark. And then, dependent on light, comes "line of sight." Via "light" and "line of sight," the notion of seeing  $x$ 's colors is linked with the notion of "looking at" it. And looking at  $x$  involves turning the eyes in the right direction, thus getting the colored thing into the line of sight. There are significant asymmetries in the logical relations between these concepts. If one sees  $x$ , one is looking at it; and if he is looking at it, his eyes are turned in its direction.<sup>2</sup> But not the other way around. His eyes may be turned in  $x$ 's direction without his looking at it. And he can be looking straight at it without seeing it.<sup>3</sup> Moreover, it can be in the direction his eyes are turned and looking, without being in his line of sight—as when the opaque wall terminates the line of sight short of the thing on the other side.

14. Someone recently argued that the concept of thinking is logically independent of the concept of brain-state (consciousness). He asserted that a computer can be said to "think" though it lacks a nervous system (consciousness). He was asked if "thinks," applied in such detachment, is the same concept as applied to a human being, and he said "Yes." Not truncated, not whittled down? "No," he said. His colleague denied this. What can be said to think *must* be said to have a brain (be conscious). How is such a dispute moderated? It is a dispute about the logical relations between concepts. What was asserted amounts to the proposition that a thinking being is only contingently a being with a brain (a conscious being). Only as a matter of contingent fact is a thinking human being a brainy being (conscious), because the concepts are not

logically tied. Really? Are they, or are they not? With what in view is a decision (if any) reached about issues like this?

15. Saying that  $S$  could be said to see  $x$  without light or without looking at it or without eyes is like saying of a non-living or non-conscious thing (computer) that it thinks.

In conclusion, let us look, from close to, at another concept hitherto overlooked in our cluster, the concept "field of vision." The crucial point here concerns its relation to the concept of "field of illumination." Suppose, as we proceed, that "in the field of illumination" is conceptually equivalent to "in the light," while "in the field of vision" is equivalent to "in sight" (in view). Investigating the relation between "in the field of vision" and "in the field of illumination" will consummate our meditation on sight and light, and further elucidate their relations to other members of our cluster.

17. These two concepts significantly interpenetrate. This is baffling to the theorists who are under the spell of the traditional scientific bifurcation of nature into the physical and the mental. For them, light is just too "physical" a phenomenon to have any logical connection with so "mental" a phenomenon as sight. Hence the easy, preliminary rejection of the notion of any "real" interpenetration of fields of vision and illumination. The notion of *anything* to which both physical and mentalistic predicates apply, or to which *neither* are *nicely* relevant, is baffling to these people. They demand reduction to just one or the other, in the final analysis. Thus, even a strong attempt like Strawson's to present as primitive the concept of an entity to which both  $M$  and  $P$  predicates apply—a person—is beginning to lose its force under the attack of a renewed physicalism more or less subtly reformulated (Nagel, Wiggins).

18. Strawson's mistake in strategy was to overlook the notion of primitive *uses* of concepts or terms in favor of the notion of primitive *terms* or concepts—a mistake to which anyone with a deductivist cast of mind is prone. Many, many terms have primitive *uses*; only a few terms get elevated to the position of primitives in this or that deductive system. It is the primitive *uses* that form the "massive central

<sup>2</sup> One can look at and see something "in a mirror" without his eyes being directed at the thing. This can be acknowledged without embarrassment for the main point.

<sup>3</sup> "Simply seeing something" itself has several weaker and stronger senses; looking-straight-at-without-seeing occurs only in the stronger senses of "simply see." In the weaker senses, to be looking at something is to see it, unbeknownst. I distinguish seven senses in another to-be-published essay.

core that has no history" (Strawson words, but not his thought). These we do not outgrow, while we do grow into and out of conceptual systems with different sets of logical primitives and their derivatives serving theoretical purposes.

19. The above aside is meant to prepare us to look at "in the field of vision" and "in the field of illumination" in their primitive *uses*, not as primitive terms. Since these expressions have an academic flavor, let us also reflect on the primitive use of their equivalents, "in sight (view)," and "in the light."

20. A number of similar things can be said about both. Here are some of them.

- (1) A source of light cannot illuminate the far side of an opaque object; a seer cannot see the far side of an opaque object. And what lies beyond or behind in that direction is "in the dark," for both light-source and the seer.
- (2) Yet, the *whole* thing is said to be "in the light" (in the light-source's field of illumination), as the whole thing is said to be in sight (in the seer's field of vision).
- (3) The source illuminates the thing at *some* distance from it, so does the seer see it at some distance from the point of view; in short, the source does not have to be up against the thing to illuminate it, nor does the seer's eyes have to be in contact with the thing to see it. Indeed, without some separation, both notions of seeing and of illuminating, curdle.
- (4) Where the thing is too far from the light-source, it is indeterminately in the field of illumination—no crisp shadows cast, etc.; as when the thing is too far from the seer's point of view, its visible properties are indeterminately seen—it is not in the crisp stereoscopic zone of the field of vision. Poor illumination, poor vision.
- (5) To be in the field of illumination is to be in the "presence" of the light-source; as being in the field of vision is to be in the presence of the seer—and to be present to him.

Here endeth the list of suggestive analogies. (There are others.) Let us see next what the suggestion is, and assess it.

What is suggested is a sort of identity of fields of vision with parts of fields of illumination, in the primitive senses of these expressions. And this in

turn suggests that it is the illuminated things that "come to light" (or "come alive") as objects of sight under the right conditions for this visual objectification of the things. In short, the suggestion is that under certain conditions, things in the field of vision are quite literally identical with the things *before* one's eyes in the field of illumination. But, since a field of vision is a sort of field of consciousness, this is to say that a field of (visual) consciousness spreads out *before* one's eyes—in the primitive sense of "eyes." It is not a brain state, nor is it a state of *mind* if that is subjective in the traditional sense which requires that its contents not be really "out there." It is a state of consciousness that is not a state of mind involving beliefs, etc. Dretske is frankly mystified by "what sort of consciousness this is" (private correspondence) that, according also to him, presents us with the same things in one world. (His own analysis *tends* to make this sort of simple visual experience into "subjective appearances" in the *old* sense, something behind the eyes, not *logically* requiring a field of illumination for its occurrence and its place.)

My concluding persuasion is in favor of an identity theory of sorts, of the field containing the things that are *at once* objects of sight (in sight) and *in the light*. But the model for such identity is more the Strawsonian sort than what the reductionist theory makes of it. I am saying that both *M* and *P* predicates, in rudimentary ("primitive") senses, apply in the characterization of this "field" and the things in it. For example, the space (field) notions of "where" or "over there" are as appropriately determined by simply looking at and seeing the thing in question (P-predicate) in this matrix situation as by, say, an architect's system of coordinates (M-predicate). But it should be noted that in such radical (root) uses, the distinction between *M* and *P* predicates is not the sharp sort of thing it becomes under the sophisticated categories of the "physical" and the "mental." Taken this latter way, their application to the field of sight and light only denatures it, and caricatures the sense of the plain talk we live by in the field. I think this has some bearing on what certain phenomenologists have recently said about "lived space" and the "lived body" (I prefer "body as subject"). They should have made more also of the notion of "lived things," in the same logical vein. But I am not sure about the bearing of what I have been saying on their view, since I find it hard to understand their idiom.

Near the beginning of this essay, I gave a sketch

of the philosophical position (phenomenalism or dualism) that spawns funsterism and is spawned by it.

Now, in conclusion, I give another picture, this time of the stance taken here as a corrective.

On the whole, I have had my say in the language of descriptive metaphysics. So it has exploited the logically primitive senses of certain terms, including "where  $x$  is." To get this sense,<sup>4</sup> and to make sense of the rest of the essay, is to understand its other key terms in a logically cognate way; that is, in their primitive senses. Among these key terms are "light" and "eyes" and "body." To talk this way is to remind one of the use of "body" as subject, not just physical object; of "eyes" as what one sees and smiles with (whose "soul" is sight), not just eyeballs; and of "light" as what enlightens people by illuminating for them the world around them (the "place" of vision), thus incorporating them in the world, portions of which thus come alive as a field of extensive "states of consciousness" that are *not states of mind*. What these presuppose is not "thinking," but a living body with eyes and a nervous system that awaken into visual experience of ambient things, under the spell of light. And this is to use "world" too in its logically primitive sense. The world in this sense, and the plain talk about it, are presupposed by theoretical sophistications, such as the distinction between the physical and the mental for what these are now worth in conceptually elaborate theories of mind and matter—and of light. It will have been noticed that the account given above is neither just physicalistic, nor just mentalistic. The "space" of the things it has portrayed is specifically neither one nor the other. This is because it deals with what is foundational—the starting point, in Aristotle's sense. Aristotle said that one must be right about starting points if one is to get anything else right. He also said, "When theories disagree with the facts of perception, they fall into contempt, and involve the truth itself in

their destruction." This assumes that one can come by some "facts of perception" without help of theories: theories that presuppose them. This essay is meant to remind theorists of some such facts, or of the milieu in which they live, move, and have their being, from first to last, as seers and men of action. The things in this environment are indeed "manifest images," but objectively "out there" for encounter in Ur-space, and for intersubjective description in plain talk—not just for hortatory expression in imperatives of action or for avowals of private sense data. They can be contemplated and, in this space, be gravid with affective values of which one becomes aware by "heeding" if not just "observing" the facts of the case. However, as a scientific theorist, one gets the (correct) impression of much to be found out by hypotheses about why certain phenomena go together—much that is, on principle, "out of sight." So one takes intellectual stock of—performs the conceptual operation of "extensive abstraction" in—the situation, thereby activating categories of thought and introducing another concept of "object" and "objective"—a notion of what the "real" massive, malleable data "actually" are in "physical space." This sophisticates the notion of "cause" and "effect," or of "making happen," having generated conceptual awareness of theoretical entities and constructs. Thinking this way, one often gets the (incorrect) impression that the things one simply sees (and thus "starts with") are themselves constructs out of private subjective sensations that in themselves (*an und fur sich*) are just a big blooming confusion. Thus experience of the ground one stubs his toe against and of the toe itself is made to look like a conceptual construction calling for further more refined "higher-order" conceptualization to "clarify" the situation. Thus one is off on the Cartesian tangent. And thus is the "reality" (thinghood) one encounters without thought volatilized into hypothetical "actuality."

University of Utah

Received May 22, 1973

<sup>4</sup> This final paragraph is excerpted from my "The Pineal Gland Up-Dated," *The Journal of Philosophy* (1970), pp. 709-710.

## VII. VULGARITY

ROBERT HOFFMAN

### I. THREE KINDS OF VULGARITY

THE several senses of "vulgar" and its cognates fall into three main groups:

- "vulgar<sub>1</sub>" = df "ordinary or customary or (of opinions) widely disseminated";
- "vulgar<sub>2</sub>" = df "commonplace or unrefined"; and
- "vulgar<sub>3</sub>" = df "of offensively mean character."

We have the first sense in "Vulgar Romans could not read the literature of Roman authors, for it was written in the Classical Latin of the aristocracy" and in "The vulgar notion that elephants fear mice is silly"; the second sense in "To vulgar judgment, wealth means success"; and the third sense in Wilde's "As long as war is regarded as wicked, it will always have its fascination: when it is looked upon as vulgar, it will cease to be popular." I shall discuss vulgarity<sub>2</sub> in detail, and the others only incidentally.

"Vulgar<sub>1</sub>" is used descriptively only, whereas the other two are not. "Vulgar<sub>1</sub>" denotes a complex relational property, viz., frequency. There is no precise frequency with which anything need occur to warrant its being vulgar<sub>1</sub>; but, despite the term's vagueness, it is wholly descriptive, denoting an entirely objective relational property. "Vulgar<sub>3</sub>" is used evaluatively. For someone to say that something is vulgar<sub>3</sub> is for him to express his con-attitude toward the thing. To say that a certain notion is vulgar<sub>1</sub> is to say what is straightforwardly true or false; and, assuming that the speaker takes appropriate care in gathering his data and doing his arithmetic, his personal values are beside the point. But if he says that the notion is vulgar<sub>3</sub>, then his personal values are relevant to at least the *prima facie* acceptability of his judgment. Whatever else he may be doing, he is plainly expressing his attitude toward the notion, and so we want to know from what values that attitude springs. Vulgarity<sub>1</sub>-judgments are purely descriptive, whereas vulgarity<sub>3</sub>-judgments are evaluative and cannot be analyzed without remainder into descriptive de-

terminants, though they can be correlated with them.

Now, vulgarity<sub>2</sub> does not have the same conceptual status as vulgarity<sub>3</sub>. Approximate synonyms of "vulgar<sub>3</sub>" are "base," "boorish," "crass," (of behavior) "crude," "obscene," and "smutty," none of which is an approximate synonym of "vulgar<sub>2</sub>." Its approximate synonyms are "uncultivated" and "undiscriminating," neither of which is an approximate synonym of "vulgar<sub>3</sub>." These synonyms indicate that even if both "vulgar<sub>3</sub>" and "vulgar<sub>2</sub>" express con-attitudes, (a) vulgarity<sub>3</sub>-judgments apply primarily to persons in virtue of their acts of *interpersonal behavior* or to the acts themselves, whereas vulgarity<sub>2</sub>-judgments apply primarily to persons in virtue of *intrapersonal performances*, such as reasoning or appraising, or to the performances themselves; and (b) vulgarity<sub>3</sub>-judgments focus the con-attitude of the judge upon *positive* properties of the thing condemned, e.g., upon *the use* of obscenity in the presence of ladies, whereas vulgarity<sub>2</sub>-judgments focus the con-attitude upon *privative* properties of the thing being condemned, e.g., upon *the absence* of appropriate subtlety in an argument or of needed sensibility in an appraisal.

There is another respect in which vulgarity<sub>2</sub>-judgments differ from vulgarity<sub>3</sub>-judgments. It is perfectly intelligible, however unusual, for someone to say of himself, "I am vulgar<sub>3</sub>." He would not be saying that he offends himself, but others; for he could well recognize and understand this fact about himself. Moreover, depending upon his values, he might want to continue to offend them or not. It is otherwise with a vulgarity<sub>2</sub>-judgment. It is not intelligible for someone to say of himself, "I am vulgar<sub>2</sub>." This feature of vulgarity<sub>2</sub>-judgments interests me, and I shall return to it. Henceforth, I shall use "vulgarity" and its cognates in the second sense only.

### II. A PARADIGM

The chief use of "vulgar" is to characterize persons and their performances of certain acts or

activities or practices, especially intrapersonal acts, activities, or practices. To say that someone or something is vulgar is (a) to express one's attitude toward the person or thing *as a whole*; (b) to express it in terms of some specific (selected) feature(s) of the person's or thing's; (c) to express a *negative* attitude; (d) to give a verdict (which is a function of rational judgment, *not* a bare evaluation); (e) to say something of special importance in getting others to see something about the person's character or the thing's nature; (f) to mention the feature(s) that uniquely or unusually individuate; (g) to make a judgment that cannot be analyzed into descriptive components, though it is coordinated with such components; and (h) to make a judgment that is apt or not.

Vulgarity-judgments are like judgments of obtuseness or dullness. Suppose that we judge someone to be obtuse. Plainly "obtuse" is pejorative, just as "acute" is commendatory. Accordingly, we are appraising. There are, of course, fit grounds for such an appraisal. If someone is habitually slow to grasp the import of things and then grasps it only superficially, we should properly characterize him as obtuse, not acute. And similarly, if a given account of something mentions only what is readily and superficially obvious to all, while not acknowledging noteworthy features of the thing, we should properly characterize the account as dull rather than acute. Acuity, whether of a mind or a particular account, is evidenced by deep penetration and subtle analysis of what is discovered. Now, profundity and subtlety are not only not precisely measurable, but are not even demonstrable to those who do not have them. In short, acuity is recognizable only by the acute, never by the obtuse. Vulgarity is subject to a similar limitation: it is recognizable only by those who are *not* vulgar.

The relationship between vulgarity<sub>1</sub> and vulgarity is fairly straightforward. "Vulgar<sub>1</sub>" describes the behavior and beliefs of the great preponderance of people in virtue of their being neither reflective enough nor sufficiently affected by training, education, or self-awareness to raise their behavior or beliefs above the dictates of prejudice. Many activities, however, depend for their legitimate effect upon the rare skill and developed capacities of those who practice them. To try to practice them without these capacities, as the vulgar<sub>1</sub> must if they try at all, results in performances that are vulgar. Now, because ordinary persons regard these activities as worthwhile, no matter why, they do try to practice them.

The question that arises here is whether or not they even understand what the activities are. Consider a paradigm—history. To the school-child and to most adults, who have not outgrown the child's belief, history is merely chronicle, i.e., a chronological record of events; to others, "history is bunk." Yet, in fact, the historian tries to discover what happened and to interpret it. To do this, he needs to know sources and competent criticism thereof, to recognize and to weigh evidence, and to understand human psychology. The layman, for example, probably would discard a known forged document, since it is not genuine; but a historian will not, for he recognizes that it still has significance, though its meaning differs from what it would be if the document were genuine. The historian tries to discover its motivations, i.e., what selfish interests it was intended to serve. In short, then, history gives us more than chronicle and is better than "bunk."

Yet people whose verdict upon some historical phenomenon is grounded in no more than their reading of some particular secondary source and who therefore fail to satisfy even the first requirement of doing historical research and analysis nevertheless volunteer "authoritative" explanations of the phenomenon, (say) the decline and fall of the Roman Empire: failure to maintain the self-governing city system; or two castes, one ever more idle, the other ever more oppressed; or emigration of its urban population to rural areas; or excessive urbanization, which antagonized the rural classes; or slavery; or dilution of the slave class by growing serfdom; and so forth. The historian's difficulty in chronicling the facts of Rome's decline and fall, and in interpreting them, does not deter the ordinary person. He produces "history" untouched by even the awareness of documentary evidence or of alternative interpretations of it, let alone by the scepticism, imaginativeness, and rigor appropriate to historical inquiry. His "history" is vulgar.

What is true of the paradigm is true of intellectual activity in general, no matter what its subject. The vulgar are in every department of life. To transcend intellectual vulgarity, one must be liberally educated. To be trained merely in a particular subject is not thereby to become able to engage non-vulgarly in any intellectual activity different from it. The task of liberal education is not to train a person to attain professional competence in every discipline of higher education, but sufficiently to familiarize him with each for him to become aware of its interrelationships and to appreciate its methods and achievements. He



learns enough not to be the fool who rushes in where angels fear to tread.

### III. VULGARITY AND ITS RELATIONS

Thinking, in the sense of merely making inferences from unanalyzed and uncritically accepted assumptions, is something all but the severely mentally defective can do. But *critical* thinking is not. To think critically, one needs (a) to have mastered language, because systematic analysis of discourse illuminates our contentions, (b) to have detailed knowledge of the subject discussed, and (c) to be conversant with other subjects that directly impinge upon the one being discussed. It was his recognition of the relative impotence and ignorance of the average human mind in action that moved James Fitzjames Stephen to write (concerning religious and moral matters) that

people should not talk about what they do not understand. No one has a right to be . . . intolerant of doctrines which he has not carefully studied . . . [M]ost people have no right to any opinions whatever upon these questions, except in so far as they are necessary for the regulation of their own affairs. [James Fitzjames Stephen, *Liberty, Equality, Fraternity*, second edition (1874) (Cambridge, 1967), p. 108.]

What he wrote of these matters applies no less to any other matter that requires subtle and informed intelligence to understand or to resolve.

Consider a simple example. Nowadays, some advocates of unrestricted abortion try to justify their belief by asserting that every woman has a moral right to do whatever she wants to do with her own body. This putative justification fails because, apart from any other consideration, *no* woman has a moral right to do whatever she wants to do with her body. Assault and battery are two immoral bodily acts. Trespass often is another. Whether or not abortion is a fourth is the question to be decided. To determine an extremely important moral issue by recourse to a slogan, but especially a slogan whose falsity is demonstrable on the very principles of morality and law affirmed by those who invoke it, is vulgar.

The vulgarity here is not mere stupidity or ignorance. It is stupidity or ignorance about a certain kind of issue and linked with a particular attitude. First, accepting the opinion has important consequences in personal or legislative behavior; secondly, to understand the subject matter in which the opinion inheres requires an informed and acute mind; thirdly, the opinion betrays consider-

able insensitivity to the kinds of issue\* that such a mind would recognize and to the methods by which it resolves them; and fourthly, in proportion to its degree, the stupidity or ignorance begets an easy omniscience about the specific issue. In this instance, vulgarity is a kind of pretentiousness that lies in the combination of these features. Another example of this kind of vulgarity is found in the second of the following verdicts:

1. Like other eminent Victorians [Mill] combined intellectual distinction with a very admirable character. [Bertrand Russell, 'John Stuart Mill' (1955).]
2. John Stuart Mill in *On Liberty* is an egotistical jerk who believed that everyone thought as he did. [A student of acting, untrained in philosophy.]
3. In the first place, no writer of the present day [other than Mill] has expressed himself upon these subjects [of *On Liberty*] with anything like the same amount either of system or of ability. In the second place, he is the only modern author who has handled the subject, with whom I agree sufficiently to differ from him profitably. Up to a certain point I should be proud to describe myself as his disciple, but there is a side of his teaching which is as repugnant as the rest is attractive to me, and this side has of late years become by far the most prominent. [James Fitzjames Stephen, *Liberty, Equality, Fraternity*, second edition (1874), ch. 1.]

What is interesting here is not the student's ignorance of the beliefs expressed in *On Liberty*, nor even the contrast between a sympathetic philosopher's and a critical jurist's respectful appraisals of Mill and the enormous stupidity of the student's remark, but primarily the pretentiousness and, in this case, also the arrogance that the ignorance and stupidity betray. Here is someone utterly lacking in the discernment and temperament appropriate to appraising Mill's mind and character who nevertheless confidently and preposterously judges both in quite extreme terms. Again, the vulgarity lies in a complex relationship discernible only by the non-vulgar.

Vulgarity, I emphasize, is not identical with pretentiousness or ignorance or stupidity or arrogance, nor with any other characteristic or combination thereof. "Vulgar" does not describe something in the sense of predicating a property whose inherence in an object is a matter of mere empirical observation. Nor is its predication logically entailed by any description of the thing of which it is predicated. To be sure, in the absence of certain descriptive properties or in the presence of others, the ascription of vulgarity would be im-

proper. To speak, for instance, of a scholar's erudite and subtle account of something, (say) Norman Kemp Smith's of Descartes' philosophy, as vulgar, would be absurd. But this fact does not require that we postulate an entailment relationship between the descriptive properties and the ascriptive characteristic, for perceiving the relationship between them requires a special awareness of the *appropriateness* of the ascription in the circumstances. This awareness, as I have already written, is possible only to those who are not vulgar.

Now, two things need to be said here. First, the ascription of vulgarity is not merely the expression of personal repugnance toward some person or thing. For the ascription is in terms of some specific selected feature(s) of the person's or thing's that characterize it unusually. And although if the great majority of people were to judge (say) Mill or Descartes, they would judge them vulgarly; still, since they do not ordinarily judge them at all, their judgments, when made, have features that characterize unusually vis-a-vis the judgments of qualified persons.

Secondly, the aforementioned features are objectively discernible and describable, though only by certain persons. This latter restriction poses no difficulty of principle, however, for tests are conceivable by which we can distinguish this class of persons. These tests are the very tests that distinguish the competent from the incompetent in a given discipline. Consider an example from literary criticism. In Milton's "On the Morning of Christ's Nativity," there are these lines:

Nor is Osiris seen  
In Memphian Grove, or Green,  
Trampling the unshowr'd Grasse with lowings loud:  
Nor can he be at rest  
Within his sacred chest,  
Naught but profoundest Hell can be his shroud,  
In vain with Timbrel'd Anthems dark  
The sable-stoled Sorcerers bear his worshipt Ark.  
  
He feels from *Juda's* Land  
The dredged Infants hand,  
The rayes of Bethlehem blind his dusky eyn . . .

[213-23]

Suppose that a reader knows none of the following: that Milton is writing in a tradition according to which the birth of Christ puts the pagan gods to rout; that (among other things) Osiris is one of those gods, that he was murdered by being locked in a chest thrown into the Nile, that he was subsequently buried at Memphis, where he arose as Apis, the sacred bull; that Christ is identified with rain:

"He shall come down like rain upon the mown grass; as showers that water the earth" (*Ps.* lxxii, 6), and with light: "I am the light of the world: he that followeth me shall not walk in darkness, but shall have the light of life" (*John* viii, 12); and that according to one tradition, the Nativity occurred in Bethlehem: "... Jesus was born in Bethlehem of Judaea ..." (*Matthew* II, i). Surely that reader's ignorance of the subject matter and background of the work incapacitates him from understanding it fully or appreciating it aesthetically, though he may nevertheless be able to grasp its tenor. Under these conditions, his judgment of the work would be vulgar.

That judgment is as clearly pointless as those of someone who discusses chemical change, but is unaware of the distinction between atoms and molecules or mixtures and compounds; or number theory, but is prepared to divide by zero; or economic adjustment of supply to demand, but has no understanding of supply potential as one of its determinants; or law, but is ignorant of such notions as strict liability and avers that proof of knowledge of wrongdoing is always necessary for conviction of a criminal offense; and so forth. In calling the judgments "pointless," I mean that not only would none help any professional in the respective subjects to understand them better, but the judgments would probably misinform or confuse persons who do not already understand the subjects.

#### IV. INTELLECTUAL COMPETENCE

Vulgarity, then, is related to lack of intellectual competence. This accords with what I wrote above about vulgarity-judgments focusing the judge's con-attitude upon *privative* properties of the thing condemned. Competence is the disposition to perform competently. A person's intellectual performance in some domain, at a given time, is *perfectly competent* if and only if, at that time, he (a) understands and uses only methods of inquiry appropriate to the subject, (b) knows everything relevant to the particular issue he considers, (c) vividly apprehends what is relevant to it, (d) recognizes the bearing upon the issue of whatever is relevant to it, and (e) is impartial, i.e., determines his beliefs about the issue and the intensity with which he holds them by the balance of evidence or reasons rather than by unconscious motives, or by internal or external constraint, or by any desire that a particular belief(s) be accepted or rejected

on non-rational grounds. Perfect intellectual competence in some domain is the disposition always to perform perfectly competently in it.

Necessarily, the foregoing definitions of "a perfectly competent intellectual performance" and of "perfect intellectual competence" are not satisfied by any actual performance or competence, for no one knows *everything* relevant to any particular issue. For some kinds of issue, future investigation invariably yields new evidence or reasons; and for others, such evidence or reasons are a constant possibility. But this should not trouble us. For in trying to understand competence, we should recognize at the outset that it admits of degree and that a person's actual competence only asymptotically approaches *perfect* competence.

It helps to distinguish a *perfectly* competent intellectual performance from a *fully* competent one. To do this, we need only to replace condition *b* by one according to which the person has the capacity to defend his solution against all criticism of it that would be raised by a perfectly rational person who is sufficiently well-informed to hold the sum total of relevant beliefs actually held by all persons when the issue actually is considered. This new concept avoids the *necessary* vacuity of its predecessor, for it is *contingent* that no such perfectly rational and well-informed person exists. Moreover, although a perfectly competent intellectual performance necessarily would yield a correct solution, a fully competent one might not, since the sum total of actually held relevant beliefs may be insufficient to determine a correct solution. Yet, although these beliefs *might* not, they probably would. This last fact is important, for it reflects an acknowledged connection between competence and correctness, namely that competent people are more likely than incompetent ones to arrive at correct solutions. Furthermore, the new emphasis on defending a solution underlines yet another acknowledged feature of intellectual competence, to wit, the capacity to defend one's contentions rationally. Those who sufficiently lack intellectual competence, i.e., who cannot defend their contentions rationally, have no intellectual right to claim truth for them or to assent to them. And, when the subject is of great intellectual or practical importance, not to be able to respond rationally to rational criticism so as to defend one's judgment or to understand why to withdraw it is to affirm it vulgarly.

But although someone *may* be fully competent in the sense specified, no one *actually* is. Some people are competent in various degrees, and other people

are incompetent. Some marks of competence, such as the capacity to make reasonable inferences, are invariant from subject to subject. Others vary with the subject: knowledge of tensor analysis, but not of mythology, is necessary to be competent at general relativity theory, whereas the reverse is true with respect to interpreting much of Milton's poetry. Moreover, some subjects, such as general relativity theory, are protected from, though they are not impregnable to, incompetents because to discuss them requires mastery of a highly technical and abstract formal language. Other subjects are not so fortunate, for they are discussed in natural languages; and vulgar people believe that there is therefore nothing in the subjects that demands more than their own capacity to talk.

Specifying criteria of competence in a given domain presupposes high competence in it. In practice, this specification is made *implicitly* by those who professionally engage in the discipline's various activities, especially in rationally criticizing each other's work. If the purpose of professionals were merely to talk or to write about their subject, there would be no need for rational criticism at all. Indeed, it would only inhibit productiveness. But mere productiveness is not what is wanted: competent work is. Rational criticism aims not only at eliminating particular errors, but at preventing trespass by incompetents and, thereby, at elevating the practice within the domain.

We can now see the relationship between vulgarity and competence. Vulgarity consists in (a) an incompetent or very low-competence performance at something of considerable intellectual or practical importance, the level of competence being so low that the performance betrays (b) the performer's non-appreciation of the subject, and (c) his belief that his performance expresses his mastery thereof. The betrayal is crucial. Incompetent performances can be given even by the highly competent, though uncharacteristically by them, and characteristically by beginners, who do not, however, claim mastery. A vulgar performance is not merely incompetent, but betrays both non-appreciation and conceit. Someone who habitually performs vulgarly is not only *being* vulgar each time he so performs, but *is* vulgar *sans réserve*. For him, a non-vulgar performance is uncharacteristic.

#### V. AESTHETIC AND EMOTIONAL VULGARITY

My explication of vulgarity has centered on intellectual vulgarity. There are, of course, other

kinds, e.g., aesthetic and emotional. A full analysis of either kind would not be to the point here, so an example of each will suffice to explain them. First, aesthetic: The vocal performances of many popular and rock-and-roll singers betray incompetent breath control, inaccurate pitch, extremely limited pitch-range, uneven tonal quality through this range, poor dynamic control, and so forth—all of which are marks of vocal incompetence. When such performances combine with ignorance of what vocal mastery is and belief that the performances express it, they are aesthetically vulgar.<sub>2</sub> Second, emotional: The general notion of friendship, in contradistinction to (say) Aristotle's or Kant's or Emerson's or Buber's, does not include all that is proper to it, viz., spiritual affinity, likeness of character, knowledge of each other's "private life," desire for each other's welfare and happiness, desire for each other's company, reciprocity of (leisure) pursuits, the special duty of ministering to those usually hidden needs that transpire in the course of coming to understand each other's "private life," and so forth. Rather, the general notion is restricted primarily to desiring each other's company, though it includes some of the others marginally. It thereby reduces friendship to mere friendly acquaintance-ship, which greatly differs from it conceptually and in value. Just as intellectual vulgarity depends upon failure to attend to what is appropriate in coming to one's beliefs, so these two kinds of vulgarity depend upon failure to attend to what is appropriate, respectively, in developing one's vocal technique and one's disposition to cultivate one's character and to share oneself with another. And in each instance, the incompetence and complacency of the performance betray that the performer is unaware of what an excellent performance is and that he takes his own to exemplify what excellence is.

Now, whatever domains of human endeavor have considerable intellectual or practical importance can have vulgar performances and performers in them. Depending upon how extensive is the range of activities over which a person performs vulgarly and how frequently his vulgar performances occur therein, he himself is or is not vulgar *sans réserve*. I do not know how to draw this line precisely, i.e., what range or what frequency to require. Nor do I believe that my ignorance here is especially burdensome. For the important thing to notice is that any sensible judgment that someone is vulgar *sans réserve* must be based upon judgments of vulgarity in particular domains, and that it is possible, however difficult, to justify a judgment of vulgarity

*sans réserve*. Like any other claim to knowledge, it is warranted only by answering all rational criticism raised against it. Whether or not this can actually be done in given particular circumstances is a contingent matter. But there is no difficulty of principle here.

## VI. HOW VULGARITY IS FOSTERED AND WHAT IT DOES

In closing, I want to discuss very briefly how vulgarity is fostered and what it does. Many expressions or concepts depend for their significance or legitimate effect upon their occurrence in (special) rare contexts. To use them in unsuitable places is to vulgarize them. The connection between our language and our thought and feeling is sufficiently intimate, I think, for degradation of language to cause degradation of the others. An example that has remained in my mind for more than a dozen years—though I can no longer recall whose example it is—will serve well here. In a beer advertisement, a man is pictured sitting in an armchair and drinking beer. On the floor are strewn newspaper pages whose headlines read, "Cities Bombed" and "Famine." A newspaper in his hand has the headline, "Invasion." Under the picture is something like "In a world of strife, where can a man replenish his courage and reaffirm his faith?" The author from whom I am borrowing emphasizes that where words such as "courage" and "faith" are used thus, the relevant concepts are lost to us; and, to that extent, we cease to understand ourselves.

Here is another telling instance of the phenomenon. During the 1973 Hollywood Academy Awards presentation, an actor intoned a speech roughly paralleling the Biblical account of the creation: "On the first day the members of the Academy offered names in nomination . . . On the sixth day there was weeping and gnashing of teeth. And on the seventh day Price Waterhouse [the ballot counter] rested." The aforementioned author, whose name I cannot recall, remarks that things of this sort do not blaspheme. For blasphemy provokes, whereas they trivialize. An object worthy of the greatest wonder is, in this instance, roughly equated with a public display of narcissism; and the careless person, unconsciously accepting this equation of the sublime with the small, gradually becomes incapable of distinguishing the one from the other. His mind and heart are diminished.

Now, examples of this sort can be drawn from life indefinitely. They typify what "the common man" feeds upon: someone of whom he has never heard declares himself a "friend" at what is really some vast, impersonal, and probably discourteously inefficient bank; the question of which soup to eat is expressed in lines imitative of Hamlet's "To be or not to be" soliloquy; the brilliant and overwhelming sonorities that begin Strauss' *Also Sprach Zarathustra* are background in a pain tablet advertisement; and so forth. Is there any surprise, then, to find that a man's "worth" has come to mean what he owns?

Many of our institutions—sometimes even our

colleges—are sufficiently ignoble not to content themselves with decently ignoring the intellectual and emotional treasures of mankind, but insist upon appropriating them for no other purpose than to reduce them to the triviality of some petty self-service. These institutions, quite usually, are those that address the great mass of men and whose vulgarity makes men indifferent, and even impervious, to the very influences that might otherwise make them human. As degradation of language begets degradation of thought and feeling, so the latter begets ignoble ideals and conduct. To deal adequately with the moral consequences of vulgarity would, however, require another paper.

*York College of the City University of New York*

*Received October 4, 1973*

## VIII. SOCRATIC DEFINITION

JOHN BEVERSLUIS

NO reader of Plato's early dialogues can fail to be struck by the centrality and the philosophical importance ascribed to definition<sup>1</sup>. Socrates is constantly asking the What-is-*X*? question, constantly looking for that character or complex of relations common to a number of instances whose presence accounts for their being, and for our calling them, *X*'s. Wittgenstein has called this tendency "the craving for generality" and holds that its correlative, "the contemptuous attitude towards the particular case,"

has shackled philosophical investigation; for it has not only led to no results, but also made the philosopher dismiss as irrelevant the concrete cases, which alone could have helped him to understand the usage of the general term. When Socrates asks the question, "what is knowledge?" he does not even regard it as a *preliminary* answer to enumerate cases of knowledge.<sup>2</sup>

Plainly, there is one sense in which such a characterization of the Socratic attitude toward the particular case cannot be faulted. For Socrates typically does reject out of hand every attempt to answer the What-is-*X*? question in terms of examples, thereby manifesting his "contempt" for them. But there is another, and no less typical, strategy which he is wont to employ in the early dialogues; and the frequency of its occurrence renders Wittgenstein's characterization significantly incomplete. That strategy consists in an appeal to particular cases as counterexamples to a proffered definition, and whose function, as such, is to provide an apparently conclusive demonstration of the necessity for amending or abandoning it. Such an appeal betrays none of the contempt of which Wittgenstein speaks; on the contrary, insofar as it implies that a compatibility with "what we would say" is a necessary condition to be satisfied by any adequate definition, it appears to confer a normative status upon the particular case.

This extraordinary ambivalence on the part of Socrates cries out for clarification. Accordingly, in

this paper I wish to examine the traditional interpretation of the Socratic Theory of Definition. I wish, on the one hand, to render epistemologically intelligible the sense in which Socrates does, in fact, dismiss the particular case, and thereby exhibit the philosophical sources of the "craving for generality" peculiar to his methodology. On the other hand, I wish to show that his appeal to particular cases is radically inconsistent with his simultaneous dismissal of them as irrelevant for answering the What-is-*X*? question. *On the hypothesis* that the traditional interpretation is a reliable explication of the texts, I shall argue that the Socratic Theory of Definition does not admit of a coherent formulation.

### I. THE WHAT-IS-*X*? QUESTION: LOGIC AND METAPHYSICS

*Euthyphro* provides a convenient point of departure. Like many interlocutors encountered by Socrates in the early dialogues, Euthyphro initially responds to the What-is-*X*? question by producing what he takes to be an example of an *X*. Having been asked what Piety is, he confidently declares: "Piety is doing as I am doing; that is to say, prosecuting any one who is guilty of murder, sacrilege, or of any similar crime . . ." (5e). As everyone knows, Socrates is never pleased with such a reply. But what exactly is it that displeases him?

It needs to be noted that Socrates' displeasure does not arise from any disagreement or doubt on his part as to whether the alleged example of *X* is, in fact, an example of it. Indeed, he typically concedes this point at once. Rather his displeasure arises from his dissatisfaction with the *kind* of answer which his question has elicited: instead of having addressed himself to the discovery of the *eidos* of Piety, Euthyphro has simply made ostensive reference to a particular action which he believes is pious. That is, he has confused definition with

<sup>1</sup> This essay is a revised version of a paper read to the Tennessee Philosophical Association Nov. 11, 1972. I wish to acknowledge my indebtedness to my commentator, Martha Osborne of the University of Tennessee, and to C. Grant Luckhardt of Georgia State University and Anthony Nemetz of the University of Georgia.

<sup>2</sup> *The Blue and Brown Books* (Oxford, 1958), pp. 19-20.

enumeration, thereby evidencing his misunderstanding of Socrates' question.

Two radically different accounts of how this question is to be elucidated co-exist uneasily in the literature. Some commentators<sup>3</sup> have supposed that, in asking the What-is-*X*? question, Socrates is interested in discovering *the meaning of a word*. According to their view, the request for an *eidos* need not be taken as carrying with it any ontological commitment; for the question: What is Piety? is reducible without remainder to: What is the meaning of "piety"?

Other commentators<sup>4</sup> bristle at this analysis. On their view, the What-is-*X*? question is to be construed neither as primarily nor even as importantly concerned with the meaning of a mere word. For it is a request for a real definition, a request, that is, for the *eidos* of the thing (*pragma*) Piety. Accordingly, the ontological character of the question is to be regarded as central and irreducible.

It is sufficient for my purposes in this paper simply to have mentioned these two quite dissimilar accounts. For I have thereby rendered harmless in advance one particular objection which might otherwise possess a certain cogency, namely, that the What-is-*X*? question is not analyzable solely in terms of a request for a definition; an objection which some philosophers would want, and surely ought, to put forth against any thesis which depends upon so controversial an analysis. My thesis, however, is such that it does not matter which of these existing accounts one accepts. For it applies to both with equal force. However the What-is-*X*? question is elucidated, whether as a logical claim about meaning or as a metaphysical claim about the *eide*, there remains a more fundamental issue involving the relation of epistemological priority between a knowledge of an *eidos* on the one hand and an ability to recognize its instances on the other. Despite their disagreement concerning the proper elucidation of the What-is-*X*? question, the two positions do not differ concerning this epistemological issue; their respective accounts are, in fact, identical. Hence I shall argue that, owing to this common epistemological thesis, the Socratic Theory

of Definition is incoherent on either account of the What-is-*X*? question.<sup>5</sup>

## II. THE EPISTEMOLOGICAL PRIORITY THESIS

Given his wish to discover the *X*-ness common to those things which are *X*, there is something logically peculiar about Socrates' characteristic rejection of particular examples as irrelevant for this purpose. For if it is the common character that he wishes to discover, how can he systematically disallow a concern with those very particulars to which it is common? Far from being irrelevant, is not such a prior gathering of instances the necessary starting-point of the inquiry?

The awkward fact is, however, that this presumably self-evident claim does not appear to have impressed Socrates himself as being self-evident at all; for his usual response to the production of an example on the part of some interlocuter is that of a rebuke followed by a restatement, and often painstaking elaboration, of his original question. But without recourse to particular cases and the ordinary meanings of words, how is the inquiry even to begin? More pointedly, what sense are we to attach to a request for that which all *X*'s have in common which includes as one of its procedural stipulations that, in answering it, we are not to take into account any *X*'s?

The Socratic rejection of examples is, therefore, problematic. It can, however, be rendered more intelligible once it is grasped that what is at issue is not simply the distinguishing of definitions from examples, but the relation of epistemological priority which, according to the traditional interpretation, Socrates believes to hold between them. At *Euthyphro* 6d-e he declares:

Tell me the nature of the idea of Piety and then (*kina*) I shall have a standard (*paradeigma*) to which I may look and by which I may measure actions . . . Then I shall be able to say that such-and-such . . . is pious and that such-and-such . . . is not.

And at *Lysis* 223b he muses:

O Menexenus and Lysis, how ridiculous that you two boys and I, an old boy, who would be one of you, should

<sup>3</sup> See, for example, R. C. Cross, "Logos and Forms in Plato," in *Studies in Plato's Metaphysics*, ed. by R. E. Allen (London, 1965), pp. 27-29.

<sup>4</sup> See R. S. Bluck, "Logos and Forms in Plato: A Reply to Professor Cross," in Allen, *op. cit.*, pp. 34 ff; Sir David Ross, *Plato's Theory of Ideas* (Oxford, 1951), p. 16; I. G. Kidd, "Plato," *The Encyclopedia of Philosophy*, ed. by Paul Edwards (New York, 1967), Vol. VII, p. 484.

<sup>5</sup> Although it is misleading to speak of the What-is-*X*? question as a request for a definition, I shall continue for two reasons to use that term: (i) the precedent (however misleading) for doing so provided by the literature; (ii) the difficulty of hitting upon a linguistic alternative which is not equally misleading, even more obscure, and stylistically tedious as well (for example, the "apprehension" of . . .).

imagine ourselves to be friends . . . as we have not yet been able to discover what a friend is.

Concerning the relation of epistemological priority in question, W. K. C. Guthrie writes:

Socrates said that you cannot discuss moral questions like how to act justly, or aesthetic questions like whether a thing is beautiful, unless you have previously decided what you mean by the concepts "justice" and "beauty" . . . Until these are fixed, so that we have a standard in our minds to which the individual actions and objects can be referred, we shall not know what we are talking about . . .<sup>6</sup>

R. E. Allen has recently championed the same thesis:

It should be noted that neither in the early dialogues nor in the middle dialogues is our knowledge of Forms derived from recognition of similarities; for it is only through the use of the Form as a standard that we may be assured that similarity in fact obtains . . . Without knowledge of [the *eidē*], the moral life becomes guess-work, and guess-work marred by mistake.<sup>7</sup>

That is, a concern with examples is not tolerated by Socrates for the following reason. For him to allow that the investigation into what *X* is may begin inductively by means of a scrutiny of known examples of *X* would be tantamount to his holding that we have *already* correctly identified them as examples. The view of epistemological priority ascribed to him, however, is precisely the opposite. That is, the ability correctly to identify something as an instance itself presupposes that we know its *eidos*.

If this interpretation is correct, it follows that Socrates must be taken as denying the apparent truism of which G. E. Moore spoke when he asserted that we may

know quite well, in one sense, what a word means, while at the same time, in another sense, we may not know what it means . . . [owing to the fact that] we are quite unable to *define* it.<sup>8</sup>

For the contradictory of Moore's claim consists precisely in the assertion that it is not the case that

someone might be competent to recognize instances of *X* (and know, in that sense, what it means) while lacking a knowledge of its definition (and not know, in that sense, what it means).

It is no part of my purpose to determine which of these views is the more philosophically plausible. My purpose is simply, given the traditional interpretation, to call attention to the *fact* of disagreement and to explore some of its implications for the Socratic Theory of Definition.

### III. THE PHILOSOPHICALLY OFFENSIVE COROLLARY

If the Socrates of the early dialogues does hold that the *eidē*, present and perceived, are themselves the sole standards by reference to which their instances can be recognized, it should hardly come as a surprise to find him disallowing appeals to instances as a preliminary to the investigation into what *X* is. For it would, of course, be self-contradictory were he to countenance such an appeal. Given this account of epistemological priority, however, Allen's conclusion (that in the absence of a knowledge of the *eidē*, the moral life becomes "guess-work marred by mistake") is too weak. For lacking such knowledge, one would be wholly ignorant not only of what *X* is, but of what things are *X* as well. And if, as the early dialogues make clear beyond all doubt, no one has such definitions, it follows as a necessary corollary that everyone is ignorant in this twofold sense.

The tradition of Platonic commentators from whose writings I have been constructing a Socratic Theory of Definition has not taken the full measure of this extraordinary inference. For it incontrovertibly commits them to the view that, for Socrates, there is a radical dichotomy between knowledge and ignorance. That is, in the pre-*Meno* dialogues there neither is nor can be any state of mind which the later Plato called "true belief": the state of mind which may justifiably be ascribed to anyone, who, though lacking a definition of *X*, is

<sup>6</sup> *A History of Greek Philosophy* (Cambridge, 1969), Vol. III, p. 352. This passage is itself an acceptably clear and representative statement of what I am calling "the traditional interpretation." However, to avoid the appearance of employing this descriptive label too loosely, as denoting an allegedly real but insufficiently identified consensus of Platonic scholarship, I offer the additional (by no means exhaustive) corroborative data: H. F. Cherniss, "The Philosophical Economy of the Theory of Ideas," in Allen, *op. cit.*, pp. 2-3; Richard Robinson, *Plato's Earlier Dialectic* (Oxford, 1953), p. 51; P. T. Geach, "Plato's Euthyphro: An Analysis and Commentary," *The Monist*, vol. 50 (1966), p. 371; R. E. Allen, *Plato's Euthyphro and the Earlier Theory of Forms* (London, 1970), pp. 72, 116; A. E. Taylor, *Plato* (New York, 1956), p. 47; I. M. Crombie, *An Examination of Plato's Doctrines* (London, 1962), p. 57; Norman Gulley, *The Philosophy of Socrates* (New York, 1968), p. 9.

<sup>7</sup> *Plato's Euthyphro*, *op. cit.*, p. 48.

<sup>8</sup> *Some Main Problems of Philosophy* (London, 1953), p. 205.



yet able to manifest a recognitional ability of its instances. For it is precisely this phenomenon which their account of epistemological priority precludes.

The later Plato had, of course, attempted to account for the occurrence of true belief, the state of mind "midway" between knowledge and ignorance, by invoking the Theory of Recollection. It was by an appeal to the *eidos* as the "dimly"-perceived standard, which this theory introduced, that he both acknowledged and attempted to account for such genuine perceptions of reality as men have. But the Theory of Recollection is wholly absent from the pre-*Meno* dialogues. While the fact of its absence is well-known, the epistemological consequence for those committed to the traditional interpretation has not, to my knowledge, been made fully explicit. It needs to be emphasized, therefore, that the foregoing account of epistemological priority, taken in conjunction with the fact that the Theory of Recollection is absent from the pre-*Meno* dialogues, jointly entail that in those dialogues there can be no consistent account of the state of mind known as "true belief." The Socratic dichotomy between knowledge and ignorance must, consequently, be sharply distinguished from the view of the later Plato, a view with which it is too often confounded.

Accordingly, the intended state to which one is to be reduced by Socratic dialectic must be taken to be that of *acknowledged* ignorance, real and total. For in the absence of both knowledge and true belief, the attempt to live the good life proves abortive at its very inception. Hence the interlocuter's presumed, but premature, confidence must be undermined; everything must be thrown into question. Indeed, the dialectic is not to be regarded as having run its full course until the interlocuter has been brought to see that he "cannot say a word" in reply to Socrates. On any other account its cutting edge is blunted, and the moral urgency attaching to its aporetic, and purportedly therapeutic, character is trivialized beyond recognition.

The early dialogues, then, must be read as demonstrating this again and again: Charmides is ignorant concerning Temperance, Euthyphro concerning Piety, Laches concerning Courage. Not even Friendship emerges intact. As outrageous and

implausible as it may seem, the Socrates of the early dialogues must be interpreted as holding that, if you cannot define Friendship, you cannot know what a friend is. Nor whether you have any. Nor how to be one. The logic of the position admits of no other alternative.

#### IV. THE CONTRADICTORY OF THE THESIS AND ITS COROLLARY

I mentioned earlier that Socrates' typical response to the production of examples on the part of some interlocuter is that of a rebuke followed by a restatement of his original question. Occasionally, however, this procedure is set aside in favor of a very different one. In *Laches*, for example, he deviates from his customary policy by allowing the enumeration of particular instances of Courage, including those courageous in war, amid perils of the sea, pain, and in overcoming their own desires; and having done so, asks for "that common quality which is the same in all these cases, and which is called 'Courage'" (191d-e).

It is passages such as this one which account for W. K. C. Guthrie's observation that

the first stage [in the Socratic quest for knowledge of *X*] is to collect instances to which it is agreed by both fellow-seekers that the name ["*X*"] can be applied. Then the collected examples . . . are examined to discover in them some common quality by virtue of which they bear that name.<sup>9</sup>

How this puzzling gloss is to be brought into harmony with Guthrie's previously cited remarks is far from clear. For one cannot simultaneously affirm (a) that known instances of *X* are to be scrutinized for the purpose of discovering the *eidos* common to them, and (b) that one cannot recognize something as an instance of *X* unless one already knows that *eidos*. Indeed, these contradictory claims exhibit in strikingly limpid form the incoherence of the traditional interpretation. According to (a), the *eidos* common to a number of instances is, in principle, discoverable inductively. Such a view presupposes, of course, that we are capable of gathering known instances of *X* while lacking a knowledge of their *eidos*. To grant this, however, is to deny the relation of epistemological priority affirmed by (b).

<sup>9</sup> *The Greek Philosophers* (New York, 1950), p. 77. See also F. M. Cornford's similarly "Aristotelian" remark: ". . . from observation of individual cases, an act of insight discerns the universal latent in [particulars] and disengages it in a generalization" *Plato's Theory of Knowledge* (London, 1950), p. 185. This despite his own earlier recognition of "Plato's break with all theories deriving knowledge by abstraction from sensible objects . . ." (p. 4), and his subsequent warning that "... no satisfactory account of Platonic Forms can be given in terms of Aristotelian logic" (p. 268).

But it is precisely that relation which is said to be constitutive of the philosophical methodology underlying the What-is-*X*? question.

The question we need to ask, therefore, is not simply how it is that the interlocuter, while lacking a knowledge of the *eidos*, can be expected to gather a number of its known instances; the real question is: how, while remaining in such a state, can he be expected to gather even one?

It will, nevertheless, be recalled that when Euthyphro responded to the question: What is Piety? with an example, Socrates rebuked him for having produced *only an example* of a pious action, instead of a definition of Piety. But, given the foregoing analysis, this is surely a curious objection on Socrates' part. For if it is the case that

- (i) one's having a definition of *X* is epistemologically prior to, and a necessary condition for, the ability to identify something as an instance of *X*.

and

- (ii) Euthyphro has not produced a definition.

then

- (iii) since there are two, and only two, states of mind with respect to *X*, knowledge and ignorance, and since, owing to his failure to produce a definition, Euthyphro's state of mind cannot be that of knowledge,

it follows that

- (iv) no one, including Socrates himself, since he, too, disclaims having such knowledge,<sup>10</sup> could possibly be in a position to determine whether Euthyphro has correctly identified an instance of a pious action or not.

Traditional commentators agree that (i), the epistemological priority thesis, is the Socratic view. Their respective elucidations of the claim that one can recognize instances (I) only if one is in possession of a definition (D), admit of a common formalization, namely,

$$I \supset D,$$

<sup>10</sup> Some readers may appeal to Socratic irony here. But this diversionary maneuver overlooks the fact that the honorific term "irony" is by no means the only possible description of the "he-knows-but-he's-not-telling" phenomenon. Other commentators have not been nearly so flattering. Richard Robinson has appraised Socrates' behavior as "insincere," as involving "persistent hypocrisy," and as evidencing "a negative and destructive spirit" (*op. cit.*, p. 22). Gregory Vlastos, too, has raised searching questions about Socrates' "limited and conditional" care for the souls of his fellows. See his "The Paradox of Socrates," in *Socrates*, ed. by Gregory Vlastos (New York, 1971), pp. 7-8, 16.

<sup>11</sup> That such a concession to Euthyphro on the part of Socrates is not a mere inadvertence of which I am taking unfair advantage is borne out by the fact that he makes similar concessions to others. See, for example, *Laches*, 191a, c-e, 193a, 196c; *Charmides*, 157d-e.

which, by a few elementary logical moves, yields

$$\sim(\sim D \& I).$$

That is, the epistemological priority thesis entails that it is not the case that any one can possess a recognitional ability without a definition. At the same time, however, Socrates grants that Euthyphro *has* identified an instance of a pious action despite the fact that he has no definition of Piety. That is, Socrates grants the truth of  $\sim D \& I$ . But in simultaneously affirming

$$(\sim D \& I) \text{ and } \sim(\sim D \& I)$$

he contradicts himself.

Socrates cannot, therefore, consistently hold that a knowledge of the *eidos* is epistemologically prior to a recognitional ability of its instances, and at the same time allow that Euthyphro, while lacking the former, has somehow managed the latter. The logic of his position entails that he simply acknowledge that neither he nor Euthyphro is in a position to know whether the alleged instance is, or is not, a genuine one. Yet, Socrates does not acknowledge this. Indeed, he allows that Euthyphro has identified an instance of a pious action.<sup>11</sup> His complaint that it is only an instance does not alter the fact that, by his own admission, it is one. Socrates claims, or at least implies, that he knows this. But, given the truth of his own theory, how can he know it?

But there is a second, and far stronger, reason for holding Socrates' procedure to be self-contradictory. This reason emerges most clearly once it is noticed that Socrates himself, while professing ignorance of some general term, is nevertheless able quite confidently to come forward with examples of it. As a counterexample to Cephalus' definition of Justice as "rendering to every man his due," Socrates produces the celebrated case of returning a weapon to its mad, though rightful, owner; a case which every one present unhesitatingly agrees is unjust. On the strength of this single counterexample Cephalus' definition is instantly rejected. But how do they know? How, that is, can every one so confidently agree that the action in question is *unjust*

while Justice remains undefined? There is a curious, and wholly unjustified, reversal of epistemological priority in evidence here, a reversal which the traditional interpretation does not, and cannot, explain.

The Theory of Definition ascribed to Socrates by that interpretation requires that a knowledge of the *eidos* is a necessary condition for the ability to recognize instances of it. An attentive examination of the proceedings, however, reveals that a compatibility with instances already recognized itself constitutes one of the criteria in terms of which the adequacy of a proffered definition is to be assessed. What this means is that a proffered definition is sometimes<sup>12</sup> rejected by virtue of its being incompatible with those things already recognized as genuine instances. But if the Socratic refutation of a proffered definition depends upon an appeal to particular cases whose function is to demonstrate its incompatibility with "what we would say," the indisputably normative character of such an appeal is sufficient to contradict any view according to which a knowledge of the *eidos* is itself prior. For surely it cannot be argued simultaneously that a

knowledge of the *eidos* is a necessary condition for the ability to recognize instances and that being compatible with instances already recognized is itself a necessary condition to be satisfied by any adequate definition. If Socrates is consistently to hold the former, he must be fully prepared to ignore discrepancies, whether apparent or real, between what *X* is and those actions which are conventionally said to be instances of *X*. Why, then, does he so often appeal to these same conventional views as a basis for rejecting a proffered definition? That is, how can part of the application of the elenchus to a definition consist in showing that it is at variance with conventionally held views about Piety and Justice, and hence unacceptable, if it is the very function of knowledge of the *eidos* as expressed by the definition to constitute the sole criterion by reference to which those same views can be truly assessed?

The following dilemma can, therefore, be constructed. Either Euthyphro, without his definition, can identify instances of Piety or he cannot. If he cannot, why does Socrates himself acknowledge that he can? If, on the other hand, he can, why does he need a definition?

Butler University

Received August 21, 1973

<sup>12</sup> Sometimes, but not always. For occasionally the criterion provided by the ordinary meanings of words is not only inexplicably ignored, but scornfully rejected. At *Republic* 420 ff we are briskly assured that the guardians are really happy, although the ordinary man (here disparagingly referred to simply as part of that undifferentiated mass known as "the multitude") would not say so. But if the ordinary man can so unproblematically be pronounced ignorant concerning the nature of happiness, how can the mere appeal to his opinion at *Republic* 331c be sufficient in itself to discredit Cephalus' definition of Justice? Why is not the same sort of appeal sufficient to discredit any definition which is incompatible with "what we would say"? How important, after all, is the compatibility requirement? And by appeal to what further criterion are we to determine when it is to be applied and when not?

## IX. FOUNDATIONS OF LOGICAL THEORY

DALE GOTTLIEB

IN its mathematical development, logic is uncontroversial. Symbolic calculi are devised, theorems proved, truth-functional and set-theoretical interpretations are defined, completeness and decidability properties are investigated, all in the spirit of scientific co-operation. Yet the lion and the lamb are not reconciled. Controversy still rages when the issue is the application of these mathematically precise systems to informal matters. Kripke's formulation of semantics for modal logic has not settled the question of the intelligibility of quantified modal sentences; the appearance of "free logics" has not ended the debate over the status of singular terms; the creation of the calculus of "relevant" entailment has not won over the adherents of classical logic. This is the question which philosophy inheres after the mathematical results are in: of what use are these beautiful structures in our attempt to formulate an adequate logical theory?

A symbolic calculus finds its usefulness in logic as a tool for identifying valid arguments. Only someone who thinks that this is the sole purpose of logical theory could be content with a calculus as his entire theory. For those of us who set our sights higher, it is important to say just what we want of our logic: what questions should be answered, what phenomena explained, by an adequate logical theory. Only then can we have a basis on which to evaluate proposed logical theories, including their mathematical tools. In Sect. II, I shall take the position that logic is essentially a part of epistemology, and in particular, part of the attempt to explicate the concept of rational belief. If this point of view is accepted, then certain debates concerning the form and the adequacy of quantification theory as a logic can be adjudicated. Two such debates are considered here. The first concerns the definition of logical truth. Seventeen years ago Strawson raised a problem for the accepted defini-

tion in terms of truth preserved under substitution.<sup>1</sup> A decade later, Quine reacted to Strawson's argument by giving up the definition and proposing a new one of his own.<sup>2</sup> An account of this transition is given in Sect. I. Quine's answer rubs one's logical instincts the wrong way, but it is only after the foundations of logical theory are laid bare in Sect. II that we can say clearly just why it is wrong and in what sense the spirit of Strawson's challenge remains unanswered. Secondly, the discussion of Sect. II is applied to the problem of the so-called "paradoxes of implication."<sup>3</sup> It is argued that for the success of logical theory in its role as a partial explication of rationality, the "paradoxes" are of little importance, and hence can be safely disregarded as an unexpected but insignificant consequence of classical logical theory.

### I

The usual definition of logical truth runs roughly like this: a sentence is logically true if it is true, and its truth is preserved by performing certain acceptable substitutions for its component expressions.<sup>4</sup> Now Strawson's objection to this definition is simply that, as defined, there are no logical truths. His argument (which I am reformulating slightly) is as follows. We naturally suppose that

(S) If John is wise, then John is wise

is logically true. Substitute any predicate you like for "is wise" and, it seems, you get another truth. But take a predicate like "is sick" which is ambiguous. In some contexts, "is sick" means roughly "is mentally ill"; in other contexts it means roughly "has a bodily disease." Now substitute "is sick" for "is wise" in (S) in such a way that it is ambiguous in the resulting sentence, the first occurrence connoting mental illness and the second bodily illness. The result is

(T) If John is sick, then John is sick

<sup>1</sup> "Propositions, Concepts, and Logical Truths," *The Philosophical Quarterly*, Vol. 7 (1957), pp. 15-25.

<sup>2</sup> Cf. his reply to Strawson in *Words and Objections*, ed. by Donald Davidson and Jaakko Hintikka (New York, 1969), pp. 322-325.

<sup>3</sup> For the latest statement of A. R. Anderson denouncing classical logic on these grounds, cf. his "An Intensional Interpretation of Truth-Values," *Mind*, vol. 81 (1972), pp. 348-371.

<sup>4</sup> For a precise formulation, cf. W. V. Quine, *Philosophy of Logic* (Englewood Cliffs, N.J., 1970), pp. 49-50.

which is *false*, if John is a physically healthy manic depressive, say. Since (T) results from (S) by an acceptable substitution, the former, though true, is not logically true. Strawson argues persuasively that by skillful use of ambiguity, we could defeat the claim to logical truth of any sentence whatever. Thus it seems that we are left with a dilemma: either we must rewrite the definition of acceptable substitution so as to take account of the possibility of ambiguous predicates, or give up the definition. The problem for Quine is that the most obvious ways of accomplishing the re-write require appeals to notions such as synonymy and ambiguity, which Quine finds unintelligible.

Quine's response to this challenge is to capitulate. He abandons the substitutional definition of logical truth in favor of a new one which will be insulated from the effects of ambiguity. This is accomplished in two steps. First, we assume that we have a language in which it is possible to identify the sentences which are instances of schemata of the first order calculus. If a sentence is an instance of a provable schema then the sentence will be called provable as well. Now the new definition of logical truth will only apply to languages which satisfy what I shall call Quine's criterion: all provable sentences are true. Secondly, the new definition itself: a sentence is logically true iff it is provable. Now (T) is a sentence which is provable and false, hence the definition of logical truth does not apply to a language containing (T). And in a language to which the definition does apply, (S), being provable, will be logically true. Strawson is answered: we have a definition which is phrased without reference to meaning and which is not subject to the effects of ambiguity.

Now let me describe an extension of Strawson's problem which neither definition of logical truth can handle. A term is said to be ambiguous if it has different meanings in different contexts. Such a shift can cause a sentence to be false, even though it is an instance of a provable schema, or a derivative by acceptable substitution from a sentence like (S) which we would like to regard as logically true. But this is only the case because the difference in meaning carries with it a difference in extension. If the difference in meaning is not accompanied by a difference in extension, then these consequences of ambiguity do not follow. For example, suppose we discovered that all mental illnesses were accompanied by bodily illness, and vice versa. Then, although "is sick" would still be ambiguous, the ambiguity would not produce a difference in ex-

tension. It follows that the ambiguity of "is sick" cannot endanger the truth of any provable sentence: if the various occurrences of a given predicate in a provable sentence have the same extension, then that sentence must be true. Thus, if we had a language in which all ambiguities were unaccompanied by shifts in extension, that language could be expected to meet Quine's criterion, and hence the definition of logical truth would apply. Would the presence of such "intensional" ambiguities have any ill effects for logical truth?

Consider (T). On the assumption that the ambiguity of "is sick" is intensional only, (T) is true. Furthermore, it is an instance of a provable schema: on both definitions it will be logically true. How satisfactory is a definition of logical truth which allows (T) to qualify for that honor?

For those who do not balk at a distinction between necessary and contingent truths, synonymy, propositions etc., our dissatisfaction with (T) as a logical truth can be stated simply: we do not expect to find contingent truths among the logical truths, and (T), with its shift in meaning, expresses a proposition which is only contingently true. But the province of logic, we feel, is confined to the realm of propositions which are necessarily true. Intensional ambiguity defeats one goal of logic, viz., to be a theory of a certain kind of a priori truth. The same point can be put another way. The definition of logical truth doubles as a definition of valid argument: an argument is valid if the hypothetical whose antecedent is the conjunction of the premisses, and whose consequent is the conclusion, is logically true. Now purely intensional ambiguities will force us to regard as valid even those arguments in which equivocations take place between the premisses and the conclusion. In our case the argument

(A) John is sick, therefore John is sick

will be valid, even if "is sick" undergoes the shift in meaning envisaged above.

The trouble with the previous paragraph is that it poses the problem for the definitions of logical truth using unintelligible terminology: my sympathies are with Quine in his rejection of synonymy, propositions, *et al.* However, the point is right in spirit. A definition of logical truth which counts (T) as logically true, and a definition of logically valid argument which takes the argument (A) to be valid, are radically different from what we expect of logic and threaten to rob it of its traditional status and significance. To press this point, we need a

description of the purposes of logical theory and the criteria which a successful logic must be expected to meet. The next section is an attempt to give such a description in terms which those of a Quinean temperament will understand.

## II

Logic is the theory of argument. Thus we expect logic to distinguish correctly between valid and invalid arguments. This requirement is to be taken in the spirit of explication. Not all our judgments of validity need to be accounted for by a theory of argument: the theory may itself give us reasons for changing our minds. Also, not all intuitions are of equal weight; we may even tolerate a theory which violates some of our judgments of validity without giving us reasons to change our minds, if it has enough attractive features and the judgments it violates are "unimportant" or "not central" to the importance of logic. Nevertheless, a clear pre-theoretic grasp of what is meant by validity, or at least what kind of a characteristic of arguments it is, will be invaluable in assessing the worth of proposed logical theories. So let us inquire what is meant by calling an argument valid.

It is often said that an argument is valid when the premisses necessitate the conclusion, or when it is impossible that the premisses be true and the conclusion false. Now there are uses of "necessitate" and "impossible" for which this is simply false, and to specify that it is *logical* necessity and impossibility that is meant is just to close a very tight circle. A middle way is to refer to linguistic necessity: it is the meanings of the terms used which make it impossible that the premisses be true and the conclusion false. The objections to this idea are many and various. The very existence of such a linguistic necessity is a hotly debated issue. Furthermore, even if such a concept of necessity is allowed, we may wonder how the meanings of terms will make the truth of the premisses and the falsity of the conclusion impossible. If, for example, the meaning of the terms are to be recorded in "meaning postulates" and the impossibility deduced from them, we are again moving in a circle. For surely we shall have to require that the deductions be accomplished via valid arguments. Finally, linguistic necessity could only be a necessary condition of validity: no one counts the argument "John is a bachelor, therefore John is unmarried" *logically* valid.

In response to this last criticism, use of the notion

of linguistic necessity in characterizing validity might be shored up by an appeal to form. Not only must the meaning of the premisses and conclusion enable the former to necessitate the latter, but this must be true for every argument of the same form. But this suggestion is ill-conceived. For, arguments do not wear their forms on their faces. Different logical theories assign arguments different forms, and even within the same theory it is common to assign different (non-equivalent) forms to the same argument. Sometimes there is debate as to which of the forms is best, as in the discussions concerning singular terms, belief and action sentences, and subjunctive conditionals. Sometimes it is simply recognized that certain sentences are aptly symbolized in a number of ways. For example, "John is six feet tall" might be rendered as " $T(J, 6)$ " or as " $H(J) = 6$ " without prejudice to the arguments in which that sentence appears. The conclusion to be drawn is that the logical form of arguments is not part of the given which logic must explain, but rather part of the machinery which logic uses in its explanations. This being the case, we cannot appeal to the notion of form in framing a neutral constraint upon logical theory.

What then can be said pre-theoretically about the concept of validity? To answer this question I think we have to turn to the *uses* of valid arguments. In what contexts does the validity of an argument make a difference, and what kind of a difference does it make? My answer, in a nutshell, is this. "Valid" is part of the vocabulary of rationality. The distinction between valid and invalid arguments is used to mark the holding of certain sets of beliefs irrational, to determine when certain beliefs give rationally sufficient grounds for others, and to determine when certain theories have been "epistemologically reduced" to others. It is in this context that we can find a non-question-begging characterization of validity. It will follow that logic, as the theory of valid argument, is essentially part of an attempt to explicate rationality, and it is success in this endeavor which is the chief criterion of the value of a particular logical theory.

Let us spell these points out in more detail. We may say that a set of beliefs is contradictory if there is a valid argument from some of the members of the set to be negation of any one of the beliefs in the set. My claim is that it is a prominent feature of the use of "valid" and "contradictory" that

- (i) To hold a contradictory set of beliefs is irrational.

Of course, (i) must be carefully explained. If the argument showing my set of beliefs contradictory is long and complicated and I have no reason to suspect its existence, in what sense is it irrational of me to hold them? It is irrational to hold a contradictory set of beliefs *in the knowledge that they are contradictory*. Once the argument comes to my attention, I must at least acknowledge that some one of my beliefs will have to be abandoned. I do not have to choose which on the spot. Pending that choice, I may indeed leave all my beliefs intact, except insofar as I must admit that not all of my beliefs are true. In this sense I cease holding the *set of beliefs*, even without giving up any particular belief.

(i) is a measure of a certain epistemological priority that logic has over other considerations. Suppose that I discover that two of my beliefs, each of which has strong independent support, are contradictory. The support of each constitutes the rational warrant to accept it. But the valid argument from the one to the negation of the other makes it impossible to accept both, so one must be dropped in spite of its independent support. In the competition of rational warrants, logic always takes precedence.

It may seem that the priority of logic expressed by (i) is too broad to distinguish it from other bodies of "necessary" truths. It is not only on logical grounds that holding a set of beliefs can be declared irrational. For example, believing both "John is 6ft. tall" and "John is not 3 + 3ft. tall" is just as surely irrational as believing both "John is 6ft. tall" and "John is not 6ft. tall." There are two replies to this objection. First, the irrationality of holding the former pair of beliefs is not established via arithmetic without an appeal to logic. Rather, we note first that that pair together with " $6 = 3 + 3$ " form a contradictory set, which (i) counsels us to give up. Since among those three " $6 = 3 + 3$ " is by far the least questionable, we concentrate the irrationality on holding the other two. Secondly, we can employ (i) directly to show the priority of logic over such other types of rational warrant of which arithmetic is a prime example. For this priority is shown in a relative fashion in general. I have a set *S* of beliefs, each with its own warrant. If logic requires that I sacrifice the set, then logic takes precedence over all the types of rational warrant which are involved in supporting the various members of *S*. Now let us consider sets of mathematical beliefs. What happens when a mathematical theory is found logically contradictory, as various set

theories have been? No matter how plausible the axioms are, some revision is demanded. The presumption is that logic has priority even over axioms of mathematics, otherwise we would find ourselves with the necessity of explaining why we do not revise our logic instead, and retain the theory.

This is not to say that our judgment of validity is inviolable. Rather, the impact of (i) is that the nature of revision in logic, and hence the need for it, is conditioned by the central use of what it is that is being revised. If a procedure of the highest epistemological priority is to be revised, the consequences will be felt throughout the organization of our knowledge. It is a deep-theoretical revision, and therefore undertaken only as a last resort, and only after a superior theory has been formulated. The various proposals for the revision of logic never take the form: the set *S* of beliefs is contradictory, but our independent warrant for each of the beliefs is so strong that we shall maintain them anyway and revise logic instead. Rather, the challenges to logic come from analyses of rationality, that is of the notions of proof, evidence, and the like.

As an example, consider the intuitionist's complaint against classical logic. One cannot be justified in asserting that there are *F*s unless one has a way to find one. Now according to the law of the excluded middle and the classical definition of the quantifiers, " $\sim(x) \sim Fx$ " implies " $(Ex)Fx$ ." And yet it seems that we can "prove" the former without "proving" the latter; that is, we can give a rationally compelling argument for the former without doing so for the latter. The deduction of a contradiction from " $(x) \sim Fx$ " will refute it, and so establish " $\sim(x) \sim Fx$ ," but may do nothing toward showing how to find an *F*. Now I am not concerned with the justice of the intuitionist's case, but rather with its form. The revision in logic is proposed so as to support his general analysis of proof for certain mathematical propositions. The challenge from quantum mechanics is similar in that it relies on a form of verificationism with respect to truth.

The second pre-theoretical feature of validity on which I want to focus is this: the premisses of a valid argument are rationally sufficient grounds for the conclusion. This means that they provide a full justification for the conclusion, assuming that they themselves are justified. Valid arguments serve to transfer justification from the premisses to the conclusion. Again, we must be careful in formulating this fact. For example, suppose someone believes the premisses and conclusion of a valid

argument, but misses the argument entirely and believes the conclusion on irrelevant grounds, or connects it to the premisses via an invalid argument. His belief in the conclusion is then not justified by his belief in the premisses. Just as in (i), we must assume that the argument is recognized for the transfer of justification to take place. So I shall put it this way:

- (ii) Whatever justifies belief in the premisses of a valid argument can, together with the argument itself, justify belief in the conclusion.

Let me immediately remove one possible objection to (ii). It does not say that whatever *supports* or is *evidence* for the premisses is also, via the argument, support or evidence for the conclusion. This would be to run afoul of a standard paradox of confirmation theory: "Grass is green" supports "Grass is green and all ravens are black," but does not support "If *A* is a raven, then *A* is black" at all. Rather, (ii) says that whatever provides *sufficient* justification for believing the premisses does the same for the conclusion, when coupled with the argument. Thus, (ii) provides a clear sense in which beliefs can be said to be epistemologically reduced to one another. A valid argument shows that there is no extra epistemological risk in adding the conclusion to a set of beliefs which contains the premisses.

The presumption of (i) and (ii) is clearly most obvious in the context of axiomatized theories. For the axiomatization of a theory to be successful, it must be *complete*, i.e., every truth of the theory must follow from the axioms by logically valid arguments alone. Once it is supposed that some of the theorems have been derived by arguments which are only enthymemes, the axioms are declared inadequate. Why is this? Because axiomatization is just the limiting case of reduction of one theory to another. One paradigm method of theoretical reduction is this: define the primitives of  $T_1$  using only the primitives of  $T_2$ , and derive via logic the truths of  $T_1$  from the axioms of  $T_2$ , and  $T_1$  is reduced to  $T_2$ . When we axiomatize a theory, the theory at large is reduced in this sense to the axioms: the primitives are the same, and the second condition is met when the axioms are complete. Now this type of reduction is supposed to carry an epistemological payload. If Frege's program for

reducing mathematics to logic in this way had been successful, we would have been entitled to conclude that "... the content expressed by true propositions of arithmetic and analysis is not something of an irreducibly mathematical character, ... but to the contrary, such propositions express truths of pure logic."<sup>5</sup> Or, in Frege's own words,

Because there are no gaps in the chains of inference, every 'axiom', every 'assumption', 'hypothesis', or whatever you wish to call it, upon which a proof is based is brought to light; and in this way we gain a basis upon which to judge the epistemological nature of the law that is proved. Of course the pronouncement is often made that arithmetic is merely a more highly developed logic; yet that remains disputable so long as transitions appear in the proofs that are not made according to acknowledged laws of logic, but seem rather to be based upon something known by intuition. Only if these transitions are split up into logically simple steps can we be persuaded that the root of the matter is logic alone.<sup>6</sup>

In a complete axiomatization, the axioms "contain the content" of their logical consequences and reveal their "epistemological nature." That is, since (i) and (ii) characterize the relation of the conclusion to the premisses of a valid argument, and since there is a valid argument from the axioms to each of the truths of the theory, we may conclude that (i) one cannot rationally believe the axioms and reject any of the truths of the theory, and (ii) whatever provides justification for the axioms does so also for the rest of the truths of the theory. If, however, the axioms are incomplete, then one may very well be able to believe the axioms and reject some of the truths with rational impunity. And, one can possess full justification for the axioms and yet not for the whole of the theory. In such a case, the epistemological interest of the axiomatization is lost, and this is the reason why completeness is so important.

This, then, is my brief sketch of some of the most central uses of the concept of valid argument. The point of this sketch for the problem at hand, namely the evaluation of theories of logic, is this. The responsibilities of a theory of logic go far beyond the mere classification of arguments as valid and invalid in approximate agreement with intuition. As with any theory, we want an explanation of the basic data. The basic datum, I submit, is the usefulness of logically valid arguments, as given by (i)

<sup>5</sup> Gottlob Frege, *The Basic Laws of Arithmetic*, tr. by Montgomery Furth (Los Angeles, 1967), p. v of the editor's introduction.

<sup>6</sup> *Ibid.*, p. 3.



and (ii). A theory of logic which provided a means of identifying valid arguments, but gave no insight into how they can play this epistemological role, is of minimal value. Furthermore, a theory of logic which identified as valid some arguments which could not play this role would certainly be incorrect.

Let me conclude this section with a methodological point. I remarked above that the definition of validity offered by any theory of logic is an attempt to *explicate* that notion. Now in explicating any rich, multi-faceted concept, it is unlikely that a single definition will be found which accords perfectly with all the uses of the concept. (Often, those uses are not all consistent with one another). This means that there must be some principle of selection which determines which aspects of the concept are crucial for the particular attempt at explication which is under way. Otherwise, criticisms of inaccuracy may be entirely misplaced.

For example, consider the various set-theoretical reductions of number theory. Suppose someone proposed a sweeping objection to all of them: according to our pre-theoretical intuitions, numbers do not have members—they simply are not sets. Hence any theory which construes them as sets is false due to its conflict with those intuitions. The answer to this objection is obvious: that particular intuition about numbers, though it undoubtedly exists, is irrelevant to the purposes of the reduction. The purpose of the reduction is to satisfy Peano's axioms (or: the true sentences of first order number theory), and also to explain how numbers can be used for counting. Any progression of entities which enables us to accomplish this will do for the numbers in the context of this kind of reduction, i.e., a reduction for this purpose. If a reduction which accomplishes this happens to violate some pre-analytic intuitions, that conflict has no impact on the success or failure of the theory.

In the case of logic, we have the problem of the so-called "paradoxes of implication." True, we do not ordinarily call an argument with contradictory premisses valid, and so we shall be suspicious of a theory of logic which counts such arguments valid. But in the absence of a statement of the purpose of the explication, i.e., of the particular pre-theoretical uses of the concept which are to be the focus of the explication, we cannot tell whether this is a crippling objection or just a peripheral detail. If we adopt (i) and (ii) as our focus, I think it is clear that the paradoxes of implication belong in the

latter group. Such arguments satisfy (i), and do not clearly violate (ii). Suppose a person's set of beliefs is shown contradictory by a "paradoxical" argument, i.e., one with contradictory premisses. Then, since the premisses are among his beliefs, his set of beliefs is contradictory and cannot be rationally maintained. And there is no *harm* in saying that whatever justifies the premisses justifies the conclusion, since the conclusion cannot be justified. The dominant feeling is that these arguments are irrelevant from an epistemological point of view. But if they are irrelevant from this point of view, and it is for the sake of the epistemological significance of valid arguments that we put forward a logical theory, then the classification of such arguments as valid or invalid is of little interest. It follows that our pre-theoretical intuitions concerning them, and the fit of our theory with those intuitions, will count very little (if at all) in evaluating the success or failure of the theory.

### III

The discussion of validity in the last section provides us with constraints on logical theory which should be understood even by those who reject hard and fast distinctions between necessary and contingent truth. These constraints are offered in the spirit of Putnam's "The Analytic and the Synthetic."<sup>7</sup> Despite the absence of a precise analytic-synthetic distinction (or because of it), the constructive task of differentiating the epistemological roles of various kinds of concepts and truths goes on. Our total theory, including our logic, is tested by experience. If it fails, revision can come anywhere. *But it cannot come to all places for the same reasons.* There are assertions which can be refuted by isolated experiments, and there are assertions of deep theory which are only to be given up when a superior deep theory is proposed. I have argued that logic is essentially part of the theory of rationality, and hence is to be revised only when doing so will contribute to a more satisfactory account of rational belief and justification. In terms of revisability, logic has priority over other disciplines, even over other deep-theoretical concerns like mathematics.

If my characterization of the epistemological role of logic is correct, then we can press the objection to Quine's redefinition of logical truth as follows. Assume that we have a language which satisfies his

<sup>7</sup> Cf. Herbert Feigl and Gilbert Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Vol. III (Minneapolis, 1962).

criterion which contains (T), namely "If John is sick, then John is sick," and in which "is sick" is intensionally ambiguous. According to Quine, (T) is logically true, and hence the argument (A) "John is sick, therefore John is sick" is valid, in spite of the ambiguity of "is sick." But the argument (A), in a context in which the ambiguity is exploited, will not function epistemologically the way (i) and (ii) require. It will not be irrational to believe the premiss and the negation of the conclusion, nor will the justification of the premiss together with the argument suffice to justify the conclusion. If a theory of logic counts such an argument valid, it violates the constraints outlined in the last section and hence is unsatisfactory. Thus the complaint of Sect. I against Quine's redefinition of logical truth is this: it misclassifies certain invalid arguments as valid, and hence cannot be used to test the accuracy of logical theory.

It seems to me that this problem, limited though it is to cases of purely intensional ambiguity, is sufficient to justify rejection of Quine's definition. However, the discussion of Sect. II puts us in a position to lodge a far deeper complaint concerning the whole philosophical perspective of logical theory that Quine's position commits him to. If the goal of logical theory is to explain the unique epistemological role of valid arguments as reflected in (i) and (ii), then the theory must include more than a symbolic calculus. Such a calculus is at best a tool for identifying the valid arguments. We shall expect a theory of logic to provide a definition of validity in terms which will permit the wanted explanation

to be given. If we possess such a definition of validity, then any proposed calculus will have to be shown accurate with respect to that definition: it will have to be proved sound and complete with respect to validity as defined. This is the only *theoretical* justification a logical calculus can have. Now Quine's redefinition of logical truth is tantamount to taking the calculus at face value. Logical truth and validity are not defined independently of the calculus in such a way as to provide an explanation of the epistemological significance of valid arguments. The definition is simply: whatever the calculus says is logically true (or valid) is so. This definition cannot contribute to the explanatory goal of logical theory. Furthermore, this position provides no theoretical justification for using this particular calculus. If we have not proved the calculus sound and complete with respect to an adequate definition of validity, how then do we defend our use of it in deciding which arguments are valid? We could appeal to agreement with intuitions concerning cases, but this is a very weak kind of justification,<sup>8</sup> and it is not clear how well the first order calculus would fare. (Remember the "paradoxes.") Thus it emerges that Quine's position<sup>9</sup> in the philosophy of logic is defeatist in the extreme. The goals of logical theory as outlined in Sect. II are completely abandoned. If this is the price of avoiding the problems of ambiguity, then Strawson's dilemma still stands: either we agree to appeal to ambiguity, synonymy, *et al.*, or quantification theory will certainly be inadequate as a theory of logic.<sup>10</sup>

*The Johns Hopkins University*

*Received August 21, 1973*

<sup>8</sup> In particular, this gives us very little reason to use the calculus as an arbiter where our intuitions are unclear or divided. We have no assurance that the calculus isolates something essential to validity generally, or just happens to coincide with validity for a certain range of arguments. However, if the calculus is sound and complete with respect to an adequate definition of validity, then when the calculus counts an argument valid we know that we have a theoretical explanation of validity for that argument.

<sup>9</sup> Quine has rarely, if ever, explicitly espoused the position which I have argued that he is committed to. However, a tendency in this direction can be found in his article with Nelson Goodman: "Steps Towards a Constructive Nominalism," *The Journal of Symbolic Logic*, vol. 12 (1947), pp. 105-122. For a criticism of this position which has affinities with my point of view, see George Berry's contribution to *Words and Objections*, *op. cit.*, pp. 243-277, esp. p. 255.

<sup>10</sup> My thanks to Larry Davis for helpful discussion.

## BOOKS RECEIVED

- Allison, Henry E., *The Kant-Eberhard Controversy* (Baltimore and London: The Johns Hopkins University Press, 1973), 186 pp. \$8.50
- Blanchette, Oliva, *For a Fundamental Social Ethic: A Philosophy of Social Change* (New York: Philosophical Library, Inc., 1973), 230 pp. \$7.50
- Capaldi, Nicholas, *The Art of Deception* (New York: Donald W. Brown, Inc., 1971), 192 pp.
- Castelli, Enrico, et al., *Demythisation et Ideologie* (Paris: Aubier, 1973), 526 pp.
- Field, Walter S., *Achieving Order from Disorder* (New York: Philosophical Library, 1973), 87 pp. \$6.00
- Gert, Bernard, *The Moral Rules* (New York: Harper Torchbooks, 1973), 246 pp. \$3.95
- Grossman, Reinhardt, *Ontological Reduction* (Bloomington and London: Indiana University Press, 1973), 215 pp. \$9.00
- Grünbaum, Adolf, *Modern Science and Zeno's Paradoxes* (London: George Allen and Unwin, Ltd., 1968), 153 pp.
- Jacobson, Jerry I., *Perspectivism in Art* (New York: Philosophical Library, 1973), 177 pp. \$8.75
- Lins, Mario, *The Philosophy of Law* (Rio de Janeiro: Livraria Freitas Bastos S/A, 1971), 164 pp.
- MacKenzie, Charles S., *Pascal's Anguish and Joy* (New York: Philosophical Library, 1973), 274 pp. \$12.50
- Melnick, Arthur, *Kant's Analogies of Experience* (Chicago and London: The University of Chicago Press, 1973), 178 pp.
- Needham, Rodney (ed.), *Right and Left* (Chicago: The University of Chicago Press, 1974), 449 pp. \$19.50
- Pangle, Thomas L., *Montesquieu's Philosophy of Liberalism* (Chicago: The University of Chicago Press, 1974), 336 pp. \$10.00
- Quinton, Anthony, *Utilitarian Ethics* (New York: St. Martin's Press, 1973), 117 pp.
- Rescher, Nicholas, *The Primacy of Practice* (Oxford: Basil Blackwell, 1973), 156 pp.
- , *Conceptual Idealism* (Oxford: Basil Blackwell, 1973), 204 pp.
- Riepe, Dale (ed.), *Phenomenology and Natural Existence* (Albany: State University of New York Press, 1973), 408 pp. \$15.00
- Saner, Hans, *Kant's Political Thought* (Chicago: The University of Chicago Press, 1973), 371 pp. \$12.75
- Schultz, Arnold, *A Theory of Consciousness* (New York: Philosophical Library, 1973), 527 pp. \$12.50
- Tuomela, Raimo, *Theoretical Concepts* (New York and Wien: Springer-Verlag, 1973), 254 pp. \$19.70

## CORRIGENDA

### MONOGRAPH NO. 8: *Studies in Modality*

1. Page 23, line 9: read "relationships" (plural).
2. Page 28, Table IV, list I, line 1: supply the missing  $\forall$ .
3. Page 29, line 12: supply  $\sim$  after  $\&$ .
4. Page 31, Table VI, the 7th entry in the E-column: supply  $\forall$  after  $\sim$ .
5. Page 33, list of Compound Modes, third entry from end: supply  $\exists$  after  $\supset$ .
6. Page 37, Table VIII, entry 12, last column: change  $\exists$  to a script E.
7. Page 42, Table X, entry 15: change  $\forall$  to  $\square$ .
8. Page 43, after line 9: supply missing line: second figure syllogism, and the fourth by converting the major, resulting in a
9. Page 46, Table XIB, line 15: change  $\forall$  to  $\square$ .
10. Page 80, line 11: read "confined" for "confronted."
11. Page 80, line 19: read "anthropomorphism" for "anthropologism."
12. Page 80, end of first paragraph: supply missing sentence: From this perspective, Kant's "Copernican Revolution" has an emphatically pre-Copernican and anthropocentric aspect: it leaves man placed in his Ptolemaic position at the center of things.

# RATIO

edited by STEPHAN KÖRNER

*Volume XVI December 1974 No. 2*

- |                |  |
|----------------|--|
| RUDOLF HALLER  | Concerning the So-called 'Münchhausen Trilemma'                        |
| MORTON WHITE   | On the Use of English 'ble'-words to Express Possibility and Necessity |
| GORDON BAKER   | Criteria: A New Foundation for Semantics                               |
| M. W. BEAL     | Essentialism and Closed Concepts                                       |
| HILLEL STEINER | The Concept of Justice   |
| JOHN HARRIS    | Leibniz and Locke on Innate Ideas                                      |

Review:

MICHAEL

WELBOURNE

*The Nature of Things* (Anthony Quinton)

Price £2.25 net

Annual Subscription £4.00 net

**Published by BASIL BLACKWELL, Oxford**

Orders to: Journals Department, 108 Cowley Road, Oxford OX4 1JF, U.K.

# The Monist

*An International Quarterly Journal of General Philosophical Inquiry*

Founded 1888 by **Edward C. Hegeler**

Editor: Eugene FREEMAN

*Editorial Board*: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

*Managing Editor*: Ann FREEMAN

*EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate with the Editor in advance for Special Instructions. Papers should be 6000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.*

## SCHEDULED PUBLICATION DATES:

- |                |            |   |
|----------------|------------|---|
| Vol. 59, No. 2 | Apr., 1975 | Philosophical Problems of Death                   |
| Vol. 59, No. 3 | July, 1975 | Language, Thought, and Reality                    |
| Vol. 59, No. 4 | Oct., 1975 | The Philosophy of Mysticism                       |
| Vol. 60, No. 1 | Jan., 1976 | Bioethics and Social Responsibility               |
| Vol. 60, No. 2 | Apr., 1976 | Philosophy and Religion in the Nineteenth Century |
| Vol. 60, No. 3 | July, 1976 | New Directions in Semantics                       |
| Vol. 60, No. 4 | Oct., 1976 | Historicism and Epistemology                      |

**Editorial Office: Box 1908, Los Gatos, California 95030**

**Business Office: Box 599, La Salle, Illinois 61301**

**SUBSCRIPTION RATES:** United States: Annual (4 issues): Institutions, \$10.00; individuals, \$7.50. Single copies: \$2.75. Foreign postage: Add 25 cents to single copy rate or \$1.00 to subscription rate.

# Journal for the Theory of Social Behaviour

Edited by P. F. SECORD (Department of Urban Studies, Queens College, New York) and  
T. MISCHEL (Department of Philosophy, SUNY at Binghamton, New York) assisted by  
R. HARRÉ (Oxford University)

## Contents

T. K. DAVENEY

Intentional Behaviour

RICHARD E. CREEL

Skinner's Copernican Revolution

JAMES R. AVERILL

An Analysis of Psychophysiological Symbolism and its Influence on Theories of Emotion

ROBERT C. SOLOMON

Freud and 'Unconscious Motivation'

FRANZ SAMELSON

History, Origin, Myth and Ideology. Comte's 'Discovery' of Social Psychology

Annual Subscription (two issues) £3.50 (\$10.50)

Single Issues £1.80 (\$5.40)

Published by **BASIL BLACKWELL**

Orders to 108 Cowley Road, Oxford OX4 1JF, England

# The Philosophical Quarterly

CONTENTS OF VOLUME 24

NO. 97

OCTOBER 1974

*Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant and Stewart McNeilly.*

*Published for the Scots Philosophical Club by the University of St Andrews.*

## ARTICLES

McTAGGART'S PARADOX AND THE  
NATURE OF TIME

Ferrel Christensen

THE IRRELEVANCE OF INTENTIONALITY  
TO PERCEPTION

H. M. Robinson

THE INDIVIDUATION OF SPEECH ACTS

Jenefer Robinson

THE UNRESOLVABILITY OF ETHICAL  
DISPUTES

J. R. Flynn

## DISCUSSIONS

THEOLOGICAL FATALISM AGAIN Paul Helm

EXTERNALISM IN ETHICS Frank Snare

TRUTH AND EVIDENCE Robert Almeder

## CRITICAL STUDY

DUMMETT ON FREGE Leslie Stevenson

## BOOK REVIEWS

## LIST OF BOOKS RECEIVED

Annual Subscriptions: Individuals £3.00 (\$8.50) Institutions £5.00 (\$14.00)

Single Issues £1.00 or \$3.00 (Institutions £1.50 or \$4.00)

for orders placed directly with the Journals Manager, Scottish Academic Press

Ltd., 25 Perth St., Edinburgh EH3 5DW, Scotland

This issue completes Volume 11 (1974) of the *American Philosophical Quarterly*. Unless you already pay by banker's order or have previously given us your standing instructions, please complete this form and deliver as instructed below:

NAME.....

ADDRESS.....

.....

.....

Signature..... DATE.....

Please enter my subscription to Volume 12 (1975) of the *American Philosophical Quarterly*. (Annual Subscription \$20 to institutions, excluding the annual monograph, \$10 to individuals including the annual monograph. Please note new subscription price.)

I enclose £...../\$

or

Debit my account

When completed, this form should be posted *with your remittance* to your usual Subscription Agency or Bookseller, or direct to:

*Journals Department*

BASIL BLACKWELL & MOTT LTD

108 COWLEY ROAD, OXFORD OX4 1JF, ENGLAND